



# Variance estimation for Sequential Monte Carlo Algorithms: a backward sampling approach

Yazid Janati, Sylvain Le Corff, Yohan Petetin

## ► To cite this version:

Yazid Janati, Sylvain Le Corff, Yohan Petetin. Variance estimation for Sequential Monte Carlo Algorithms: a backward sampling approach. 2022. hal-03630333

**HAL Id: hal-03630333**

**<https://hal.archives-ouvertes.fr/hal-03630333>**

Preprint submitted on 5 Apr 2022

**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.

# VARIANCE ESTIMATION FOR SEQUENTIAL MONTE CARLO ALGORITHMS: A BACKWARD SAMPLING APPROACH

YAZID JANATI EL IDRISSI, SYLVAIN LE CORFF, YOHAN PETETIN

**ABSTRACT.** In this paper, we consider the problem of online asymptotic variance estimation for particle filtering and smoothing. Current solutions for the particle filter rely on the particle genealogy and are either unstable or hard to tune in practice. We propose to mitigate these limitations by introducing a new estimator of the asymptotic variance based on the so called backward weights. The resulting estimator is weakly consistent and trades computational cost for more stability and reduced variance. We also propose a more computationally efficient estimator inspired by the *PaRIS* algorithm of [33]. As an application, particle smoothing is considered and an estimator of the asymptotic variance of the Forward Filtering Backward Smoothing estimator applied to additive functionals is provided.

## CONTENTS

1. Introduction	2
2. Notations	3
3. Sequential Monte Carlo	4
3.1. Definitions	4
3.2. Particle filter	5
3.3. Asymptotic variance estimation in particle filters	6
4. Variance estimation with backward sampling	7
4.1. Term by term variance estimator	7
4.2. Computation for $b = 0$ and $b = e_s$	11
4.3. Variance estimators with reduced computational cost	12
4.4. A PaRIS variance estimator	13
5. Application to the FFBS	15
5.1. FFBS algorithm	15
5.2. Asymptotic variance estimator	16
5.3. Algorithm for marginal smoothing	17
6. Numerical simulations	18
6.1. Asymptotic variance of the predictor	19
6.2. Asymptotic variance of the smoother	22
References	23
Appendix A. Proofs	26
A.1. Preliminaries	26
A.2. Proof of Proposition 4.3	26

---

SAMOVAR, TELECOM SUDPARIS. INSTITUT POLYTECHNIQUE DE PARIS, 91120 PALAISEAU  
FRANCE

*E-mail address:* yazid.janati\_elidrissi@telecom-sudparis.eu,  
sylvain.le\_corff@telecom-sudparis.eu, yohan.petetin@telecom-sudparis.eu.

A.3.	Proof of Proposition 5.1	29
A.4.	Proof of Theorem 4.4	29
A.5.	Proof of Theorem 4.5	35
A.6.	Proof of Theorem 4.7	35
A.7.	Proof of Theorem 4.8	37
A.8.	Proof of Theorem 5.2	37
A.9.	Supporting results for Theorem 4.4	42
A.10.	Supporting results for Theorem 4.7	50
Appendix B.	Further algorithmic details	65
B.1.	Variance estimators for the predictor and filter	65
B.2.	GT term by term estimator of the asymptotic variance	67
Appendix C.	Technical results	67
Appendix D.	Asymptotic variance of the joint predictive distribution	68

## 1. INTRODUCTION

Sequential Monte Carlo (SMC) methods offer a flexible framework for the approximation of posterior distributions in the context of Bayesian inference, for instance in *Hidden Markov Models* (HMM). These models presuppose that the observations are defined using an unobserved process assumed to be a Markov chain. In such a setting, we are particularly interested in estimating the law of a hidden state given all past observations referred to as the filtering distribution and the laws of sequences of states given past and future observations, referred to as smoothing distributions. These distributions can be approximated by weighted empirical measures associated with random samples, usually known as *particles*. All SMC methods are based on successive importance sampling and resampling steps. When a new observation is available, new particles are sampled according to an importance distribution and then they are weighted to match the target distribution. Finally, through a resampling scheme, particles with large weights are duplicated while low weighted particles are discarded. This general procedure has been used in a wide range of applications such as signal processing, target tracking, econometrics, biology, see [3, 6, 16] and the references therein.

The quantification of the Monte Carlo error of SMC estimators is a major challenge. For a variety of SMC methods such as the *bootstrap filter* [23] or the *Forward Filtering Backward Smoothing* (FFBS) [37] algorithms, *Central Limit Theorems* (CLT) with theoretical expressions of the asymptotic variance (in the number of particles) have been derived [13, 5, 26, 15]. However, these expressions are not computable in practice and give rise to the natural question of their estimation. Since this problem appears in an online context, a critical constraint is that the samples produced by the original SMC algorithm should be recycled to compute such estimates.

This problem has received some satisfactory solutions recently: a simple estimator of the asymptotic variance has been proposed in [4] for the bootstrap filter algorithm and has since been refined in [27, 32, 18]. The computation of the associated asymptotic variance estimator at time  $t$  is based on tracing the genealogy of each particle down to time 0. Although it has been shown to be consistent as the number of particles  $N$  grows to infinity, it is particularly prone to instability when  $t$  is large, as the successive resampling steps lead to the well known path degeneracy issue. After a few time steps, particles are likely to share the same ancestor at time 0 which in turn makes the asymptotic variance estimates collapse. To overcome this degeneracy issue, [32] proposed to only trace a part of the genealogy of each particle according to a fixed-lag parameter following the fixed-lag smoothing

approach introduced in [31]. As long as the number of particles is balanced with the chosen lag, the bias introduced by considering only the most recent ancestors of each particle can be controlled as shown in [32]. However, while this alternative estimator remains easily computable, choosing an optimal lag is a non trivial task which makes this approach hard to tune in practice. We thus address the limitations of current asymptotic variance estimators in this paper.

The first contribution of this paper is to propose a parameter free estimator of the asymptotic variance associated with the bootstrap particle filter that trades computational cost for stability and reduced variance. The construction of our estimator starts from the observation that the aforementioned degeneracy is similar to that of classical SMC-based smoothing algorithms [17, 20, 35] or Particle Markov Chain Monte Carlo algorithms such as the Particle Gibbs sampler [1]. In both cases, a backward sampling step which aims at diversifying the particle trajectories has shown to be a reliable workaround that decreases the (theoretical) variance of the estimators at the expense of higher computational cost [22, 15, 33, 11, 29, 7]. We thus aim at introducing such a mechanism in the estimation of the asymptotic variance. The construction of our estimator relies on the analysis conducted in [27] in which it is shown that the estimator of [4] can be interpreted as a conditional expectation with respect to the indices that retrace the genealogy of the particles, given all the particles and ancestors. We show that this construction still holds when the distribution of the indices relies on the backward importance weights. The resulting estimator is computed by averaging auxiliary statistics that are very similar to those of the *forward implementation* of the FFBS for additive functionals [12] and can be thus updated online. The time complexity per update of our estimator is of order  $N^3$ . Driven by the efficient implementation of the FFBS for additive functionals developed in [33], we show that the computational cost of our estimator can be reduced from  $\mathcal{O}(N^3)$  to  $\mathcal{O}(N^2)$  by means of additional Monte Carlo simulation while remaining as competitive in terms of bias and variance.

We next focus on the FFBS algorithm for the estimation of smoothing estimators. Despite the fact that a CLT has been obtained for estimators based on the FFBS, no variance estimator has been proposed in the literature. We show that our previous construction enables us to fill this gap and we thus provide a consistent estimator in the case of additive functionals which are particularly critical, for instance in the Expectation Maximization framework. Again, this estimator can be computed online and in the particular case of marginal smoothing its computational cost can be drastically reduced.

The paper is organized as follows. In Section 3 we briefly review the SMC framework and discuss the current estimators of the asymptotic variance proposed in [4, 27, 32]. In Section 4, we introduce our estimator based on the backward weights, propose an online implementation and establish its asymptotic properties. In Section 5, we extend our derivations to the FFBS algorithm and provide an asymptotic variance estimator. We finally validate our results with numerical experiments in Section 6. All additional proofs and discussions can be found in the appendix.

## 2. NOTATIONS

For any measurable space  $(E, \mathcal{E})$ , we denote by  $\mathbb{F}(\mathcal{E})$  the set of  $\mathbb{R}$  valued,  $\mathcal{E}$ -measurable functions, by  $\mathbb{F}_b(\mathcal{E})$  the subset of bounded functions on  $E$ , and by  $\mathbb{M}(E)$  the set of measures on  $E$ . For any  $\mu \in \mathbb{M}(E)$  and  $h \in \mathbb{F}(\mathcal{E})$ , we write

$$\mu(h) := \int_E h(x) \mu(dx) .$$

For any transition kernel  $M$  from  $(E, \mathcal{E})$  to another measurable space  $(G, \mathcal{G})$ , define

$$M[h](x) := \int_G M(x, dy)h(y), \quad \forall h \in \mathbb{F}(\mathcal{G}), \quad \forall x \in E.$$

and write  $\mu M$  the measure defined on  $G$  by

$$\mu M(A) := \int_E \mu(dx)M(x, A) \quad \forall A \in \mathcal{G}.$$

If  $M_1$  is a transition kernel from  $(E, \mathcal{E})$  to  $(G, \mathcal{G})$  and  $M_2$  a transition kernel from  $(G, \mathcal{G})$  to  $(H, \mathcal{H})$ , then  $M_1 M_2$  is a transition kernel from  $(E, \mathcal{E})$  to  $(H, \mathcal{H})$  defined by

$$M_1 M_2(x, A) = \int_G M_1(x, dy)M_2(y, A) \quad \forall A \in \mathcal{H}.$$

In addition,  $M_1 \otimes M_2$  is a transition kernel from  $(E, \mathcal{E})$  to  $(G \times H, \mathcal{G} \otimes \mathcal{H})$  defined by

$$M_1 \otimes M_2(x, A) := \int \mathbb{1}_A(y, z)M_1(x, dy)M_2(y, dz) \quad \forall x \in E, \quad A \in \mathcal{G} \otimes \mathcal{H}.$$

In particular, for all  $N \geq 1$ , we will write  $M^{\otimes N}$  for  $\bigotimes_{i=1}^N M$ . For two  $\mathcal{E}$ -measurable functions  $f, g$  the tensor product is defined as

$$f \otimes g : E^2 \ni (x, y) \mapsto f(x)g(y).$$

The  $\mathbf{L}_q$  norm of a random variable  $X$  is  $\|X\|_q := \mathbb{E}[X^q]^{1/q}$ . The supremum norm of  $f \in \mathbb{F}_b(E)$  is denoted by  $\|f\|_\infty$ . The unit function  $\mathbf{1}$  is such that  $\mathbf{1}(x) = 1$  for all  $x \in E$ . The transpose of a matrix  $A$  is denoted  $A^\top$ . If  $A, B \in \mathbb{R}^{M \times N}$  are two matrices, then the Hadamard product  $A \odot B$  is the element-wise product, i.e for all  $1 \leq i \leq M$  and  $1 \leq j \leq N$ ,  $(A \odot B)_{i,j} = a_{i,j}b_{i,j}$ . If  $A \in \mathbb{R}^{N \times N}$ , then  $\text{Diag}(A)$  is the  $N \times N$  diagonal matrix such that for all  $1 \leq i \leq N$ ,  $\text{Diag}(A)_{i,i} = A_{i,i}$  and if  $\mathbf{x} \in \mathbb{R}^{N \times 1}$  then  $\text{Diag}(\mathbf{x})$  is the  $N \times N$  diagonal matrix such that  $\text{Diag}(\mathbf{x})_{i,i} = x_i$ . For  $(a, b) \in \mathbb{N}^2$ ,  $[a : b] := \mathbb{N} \cap [a, b]$  and  $[b] := [1 : b]$ . If  $f$  is a mapping from  $[N]^2$  to  $\mathbb{R}$ , we denote by  $\mathbf{f}$  the associated  $N \times N$  matrix such that  $\mathbf{f}_{i,j} = f(i, j)$ . Finally, we adopt the following conventions. Given some set  $\{\xi_s^i\}_{s \in [0:t], i \in [N]}$ , we write  $\xi_t^{1:N} := (\xi_t^1, \dots, \xi_t^N)$ ,  $\xi_{0:t}^{k_0:t} := (\xi_0^{k_0}, \dots, \xi_t^{k_t})$ ,  $\xi_{0:t}^{1:N} := \{\xi_{0:t}^{k_{0:t}}\}_{k_{0:t} \in [N]^{t+1}}$ .

### 3. SEQUENTIAL MONTE CARLO

In this section, we first review the bootstrap particle filter methodology and recall the main asymptotic results associated with the estimates produced by this algorithm. A state of the art and a discussion of the estimation of the asymptotic variances related to this algorithm are also presented.

**3.1. Definitions.** Let  $(X, \mathcal{X})$  be a general measurable space. Let  $M_0$  and  $(M_t)_{t \in \mathbb{N}}$  be a probability measure on  $(X, \mathcal{X})$  and a sequence of Markov transition kernels on  $X \times \mathcal{X}$ , respectively. Consider also a family  $(g_t)_{t \in \mathbb{N}}$  of non-negative  $\mathcal{X}$ -measurable functions, referred to as potentials. Throughout this paper, we make the following assumptions on  $\{M_t\}_{t \in \mathbb{N}}$  and  $\{g_t\}_{t \in \mathbb{N}}$ .

- (A1) The probability measure  $M_0$  admits  $m_0$  as probability density with respect to some reference measure  $\nu \in \mathbb{M}(\mathcal{X})$ . For all  $t \in \mathbb{N}$  and  $x_t \in X$ ,  $M_{t+1}(x_t, \cdot)$  admits  $m_{t+1}(x_t, \cdot)$  as probability density with respect to  $\nu$ .
- (A2) There exists a constant  $G_\infty > 0$  such that for all  $t \in \mathbb{N}$  and  $x \in X$ ,  $0 < g_t(x) \leq G_\infty$ .

Define the sequence of unnormalized transition kernels  $(\mathbf{Q}_{t+1})_{t \in \mathbb{N}}$  where, for all  $t \in \mathbb{N}$ ,  $x_t \in \mathbf{X}$  and  $A \in \mathcal{X}$ ,

$$\mathbf{Q}_{t+1}(x_t, A) := g_t(x_t) \mathbf{M}_{t+1}(x_t, A) ,$$

and, for any  $s, t \in \mathbb{N}^2$ ,

$$\mathbf{Q}_{s:t} := \begin{cases} \mathbf{Q}_s \otimes \cdots \otimes \mathbf{Q}_t & \text{if } s \leq t , \\ \text{Id} & \text{otherwise} . \end{cases}$$

Define recursively the sequence of measures  $(\gamma_{0:t})_{t \in \mathbb{N}}$ ,

$$\gamma_0(dx_0) := \mathbf{M}_0(dx_0) , \quad \gamma_{0:t}(dx_{0:t}) := \gamma_{0:t-1}(dx_{0:t-1}) \mathbf{Q}_t(x_{t-1}, dx_t) , \quad (3.1)$$

and let  $\gamma_t(dx_t) = \int_{\mathbf{X}^t} \gamma_{0:t}(dx_{0:t})$ . Sequential Monte Carlo algorithms aim at solving recursively the filtering problem, i.e. at estimating the sequence of probability measures defined as

$$\eta_t(dx_t) := \gamma_t^{-1}(\mathbf{1}) \gamma_t(dx_t), \quad \phi_t(dx_t) := g_t(x_t) \eta_t(dx_t) / \eta_t(g_t) , \quad (3.2)$$

respectively called the *predictive* and *filtering* measures. Note that  $\eta_t$  can be computed recursively using

$$\eta_t(dx_t) = \int \phi_{t-1}(dx_{t-1}) \mathbf{M}_t(x_{t-1}, dx_t) . \quad (3.3)$$

We motivate these definitions with the following example.

**Example 3.1. Hidden Markov models** consist of an unobserved state process  $\{X_t\}_{t \in \mathbb{N}}$  and observations  $\{Y_t\}_{t \in \mathbb{N}}$ . They respectively evolve in two general measurable spaces  $(\mathbf{X}, \mathcal{X})$  and  $(\mathbf{Y}, \mathcal{Y})$ . It is assumed that  $\{X_t\}_{t \in \mathbb{N}}$  is a Markov chain with transition kernels  $(\mathbf{M}_{t+1})_{t \in \mathbb{N}}$  and initial distribution  $\mathbf{M}_0$ . Given the states  $\{X_t\}_{t \in \mathbb{N}}$ , the observations  $\{Y_t\}_{t \in \mathbb{N}}$  are independent and for all  $t \in \mathbb{N}$ , the conditional distribution of the observation  $Y_t$  only depends on the current state  $X_t$ . This distribution is written  $G_t(X_t, \cdot)$  and admits the potential  $g_t(x_t, \cdot)$  as density (the dependence in  $Y_t$  is made implicit and we drop the second argument). Given an observation record  $Y_{0:t}$ , the predictive and filtering distributions (3.2) are then the distributions of  $X_t$  given  $Y_{0:t-1}$  and  $X_t$  given  $Y_{0:t}$  respectively.

These two distributions are of considerable interest in Bayesian filtering as they enable the estimation of the hidden states through the observed data record. Unfortunately only in a few cases, such as discrete state spaces or linear and Gaussian HMM, can they be obtained in closed form, see [3, 6] for a complete overview.

**3.2. Particle filter.** We now illustrate how to obtain empirical estimates of  $\eta_t$  and  $\phi_t$  in an online manner through Monte Carlo simulation. Assume that at time  $t$  the empirical measure  $\eta_t^N(dx_t) := N^{-1} \sum_{i=1}^N \delta_{\xi_t^i}(dx_t)$  based on random samples  $\{\xi_t^i\}_{1 \leq i \leq N}$  approximates  $\eta_t(dx_t)$ . Plugging  $\eta_t^N$  in (3.2) provides an approximation of  $\phi_t(dx_t)$ ,

$$\phi_t^N(dx_t) := \sum_{i=1}^N \mathcal{W}_t^i \delta_{\xi_t^i}(dx_t) ,$$

where  $\mathcal{W}_t^i := \Omega_t^{-1} \omega_t^i$ ,  $\omega_t^i := g_t(\xi_t^i)$  and  $\Omega_t := \sum_{i=1}^N \omega_t^i$ . Replacing  $\phi_t$  by  $\phi_t^N$  in (3.3), we obtain the mixture  $\phi_t^N \mathbf{M}_{t+1}$  which allows to construct  $\eta_{t+1}^N$  by drawing  $N$  samples from it. We first sample for all  $1 \leq i \leq N$  an ancestor index  $A_t^i \sim \text{Categorical}(\mathcal{W}_t^{1:N})$ , and then sample  $\xi_{t+1}^i \sim \mathbf{M}_{t+1}(\xi_t^{A_t^i}, \cdot)$ . The algorithm is initialized with the approximation of  $\eta_0 = \mathbf{M}_0$ ,  $\eta_0^N := N^{-1} \sum_{i=1}^N \delta_{\xi_0^i}(dx_t)$  where  $\xi_0^{1:N} \stackrel{\text{iid}}{\sim} \mathbf{M}_0^{\otimes N}$  and coincides with the *bootstrap* algorithm [23]. Note that this mechanism has been

extended in many directions in the past decades [34, 14, 6]

Alongside  $\eta_t^N$  and  $\phi_t^N$ , the particle approximation of the unnormalized marginal  $\gamma_t(dx_t)$  is given by

$$\gamma_t^N(dx_t) := \left\{ \prod_{s=0}^{t-1} N^{-1} \Omega_s \right\} \eta_t^N(dx_t), \quad \forall t > 0, \quad (3.4)$$

and  $\gamma_0^N(dx_0) = \eta_0^N(dx_0)$ . In particular,  $\gamma_t^N(h)$  is unbiased for any measurable function  $h$  [9].

In the remainder of this paper, we denote by  $\mathcal{F}_t^N$  the  $\sigma$ -field containing all the particles and ancestors up to time  $t$ , i.e.  $\mathcal{F}_t^N := \sigma(\xi_{0:t}^{1:N}, \mathbf{A}_{0:t-1}^{1:N})$ .

**3.3. Asymptotic variance estimation in particle filters.** The particle filter described above yields consistent estimators, see for instance [3, 16, 6, 30, 9] and references therein for a complete overview. Indeed, for a test function  $h \in \mathbb{F}_b(\mathcal{X})$  and under assumption (A2), the SMC estimators satisfy a Strong Law of Large Numbers when the number of particles  $N$  goes to infinity, i.e.

$$\gamma_t^N(h) \xrightarrow[N \rightarrow \infty]{a.s.} \gamma_t(h), \quad \eta_t^N(h) \xrightarrow[N \rightarrow \infty]{a.s.} \eta_t(h), \quad \phi_t^N(h) \xrightarrow[N \rightarrow \infty]{a.s.} \phi_t(h). \quad (3.5)$$

Under the same assumptions, CLTs for  $\gamma_t^N(h)$ ,  $\eta_t^N(h)$  and  $\phi_t^N(h)$  are also available [13, 5]:

$$\begin{cases} \sqrt{N}(\gamma_t^N(h) - \gamma_t(h)) & \xrightarrow[N \rightarrow \infty]{} \mathcal{N}(0, \mathcal{V}_{\gamma,t}^\infty(h)), \\ \sqrt{N}(\eta_t^N(h) - \eta_t(h)) & \xrightarrow[N \rightarrow \infty]{} \mathcal{N}(0, \mathcal{V}_{\eta,t}^\infty(h)), \\ \sqrt{N}(\phi_t^N(h) - \phi_t(h)) & \xrightarrow[N \rightarrow \infty]{} \mathcal{N}(0, \mathcal{V}_{\phi,t}^\infty(h)), \end{cases} \quad (3.6)$$

where  $\Rightarrow$  denotes weak convergence and

$$\mathcal{V}_{\gamma,t}^\infty(h) = \sum_{s=0}^t \{ \gamma_s(\mathbf{1}) \gamma_s(\mathbf{Q}_{s+1:t}[h]^2) - \gamma_t(h)^2 \}, \quad (3.7)$$

$$\mathcal{V}_{\eta,t}^\infty(h) = \sum_{s=0}^t \frac{\gamma_s(\mathbf{1}) \gamma_s(\mathbf{Q}_{s+1:t}[h - \eta_t(h)]^2)}{\gamma_t(\mathbf{1})^2}, \quad (3.8)$$

$$\mathcal{V}_{\phi,t}^\infty(h) = \sum_{s=0}^t \frac{\gamma_s(\mathbf{1}) \gamma_s(\mathbf{Q}_{s+1:t}[g_t\{h - \phi_t(h)\}]^2)}{\gamma_{t+1}(\mathbf{1})^2}. \quad (3.9)$$

An intuitive derivation of (3.7) is proposed in Section D of the appendix. The authors of [4] propose to estimate (3.7) online using the samples produced by the particle filter described above. Their estimator is based on the genealogy of the particle system induced by the successive resampling steps of the particle filter. From the indices  $A_t^i$ , it is possible to trace back the ancestors of each particle and deduce the corresponding ancestor at time  $t = 0$ . More interestingly, these ancestors can be computed in a forward way by introducing the Eve indices  $E_{t,0}^i$ . For all  $i \in [1 : N]$ ,  $E_{t,0}^i$  describes the index of the ancestor at time 0 of particle  $\xi_t^i$  and can be computed from

$$E_{t,0}^i = E_{t-1,0}^{A_{t-1}^i} \mathbb{1}_{t>0} + i \mathbb{1}_{t=0}. \quad (3.10)$$

As explained in [32], the asymptotic variance estimator of  $\eta_t^N(h)$  obtained in [4] can be written as

$$\mathcal{V}_{\eta,t}^N(h) := -N^{-1} \sum_{i,j \in [N]^2} \mathbb{1}_{E_{t,0}^i \neq E_{t,0}^j} \{h(\xi_t^i) - \eta_t^N(h)\} \{h(\xi_t^j) - \eta_t^N(h)\}. \quad (3.11)$$

We sometimes refer to  $\mathcal{V}_{\eta,t}^N(h)$  as the CLE (Chan & Lai Estimator). Note that it can be computed online in a remarkably simple way since the Eve indices (3.10) are computed recursively. However,

the counterpart of its computational simplicity is that it degenerates as soon as the ancestral paths coalesce. Indeed, it is widely known in the SMC literature that all lineages eventually end up with the same ancestor when  $t$  is large enough with respect to the number of samples  $N$  (see e.g. [20, Section 2.2] for a more detailed explanation). This means that for a fixed  $N$ , as  $t$  grows and  $s \ll t$ ,  $E_{s,0}^i = E_{s,0}^j$  for all  $(i, j) \in \mathbb{N}^2$  and  $\mathcal{V}_{\eta,t}^N(h) = 0$  for any test function  $h$ .

The degeneracy problem concerning (3.11) is partially addressed in [32] by truncating the genealogy of the particle system. Denoting  $\lambda \in [t]$  the lag and  $E_{t,t-\lambda}^i$  the ancestor of  $\xi_t^i$  at time  $t - \lambda$ , their estimator reads

$$\mathcal{V}_{\eta,t}^{N,\lambda}(h) := -N^{-1} \sum_{i,j \in [N]^2} \mathbb{1}_{E_{t,t-\lambda}^i \neq E_{t,t-\lambda}^j} \{h(\xi_t^i) - \eta_t^N(h)\} \{h(\xi_t^j) - \eta_t^N(h)\}. \quad (3.12)$$

In the regime where (3.11) degenerates, (3.12) can be made stable provided that the lag  $\lambda$  is chosen such that there is little asymptotic bias. However, besides the strong mixing case for which the authors propose a heuristic, the practical choice of such a  $\lambda$ , although crucial, is a non trivial task.

In [27] the CLE is revisited using different techniques based on [8, 1, 2]. These tools enable them to derive a weakly consistent term by term estimator of (3.7) based on the unbiased estimation of  $\gamma_t(h)^2$  and of each  $\gamma_s(\mathbf{1})\gamma_s(\mathbf{Q}_{s+1:t}[h]^2)$ , for all  $s \in [0 : t]$ . The construction of this second estimator is appealing and insightful in that it helps identifying the deep root of the degeneracy in the CLE. They indeed show that (3.11) and each  $\gamma_s(\mathbf{1})\gamma_s(\mathbf{Q}_{s+1:t}[h]^2)$  can be interpreted as a conditional expectation with respect to particle indices that retrace the ancestral paths. Indeed, by introducing discrete random variables  $K_{0:t}^1$  and  $K_{0:t}^2$  such that conditionally on  $\mathcal{F}_t^N$ ,  $K_t^1$  and  $K_t^2$  are distributed uniformly on  $[N]$  and such that for any  $s \in [0 : t - 1]$ ,

$$K_s^1 = A_s^{K_{s+1}^1}, \quad K_s^2 = \mathbb{1}_{K_{s+1}^1 \neq K_{s+1}^2} A_s^{K_{s+1}^2} + \mathbb{1}_{K_{s+1}^1 = K_{s+1}^2} C_s,$$

where  $C_s \sim \text{Categorical}(\mathcal{W}_s^{1:N})$ , then for example

$$\mathcal{V}_{\eta,t}^N(h) = -N\mathbb{E} \left[ \prod_{s=0}^t \mathbb{1}_{K_s^1 \neq K_s^2} \{h(\xi_t^{K_s^1}) - \eta_t^N(h)\} \{h(\xi_t^{K_s^2}) - \eta_t^N(h)\} \middle| \mathcal{F}_t^N \right].$$

An intuitive extension is to replace the deterministic assignments  $K_s^1 = A_s^{K_{s+1}^1}$  and  $K_s^2 = A_s^{K_{s+1}^2}$  by random ones based on backward sampling. Essentially, the backward kernel samples at time  $s$  an index  $i$  with probability proportional to  $\omega_s^i m_{s+1}(\xi_s^i, \xi_{s+1}^{K_{s+1}^1})$  (resp.  $\omega_s^i m_{s+1}(\xi_s^i, \xi_{s+1}^{K_{s+1}^2})$ ) and thus allows to consider relevant trajectories which are not necessarily ancestral trajectories.

#### 4. VARIANCE ESTIMATION WITH BACKWARD SAMPLING

In this section we present three variance estimators for the bootstrap particle filter. In Section 4.1, we lay out our methodology and derive a term by term variance estimator; its computation is detailed in Section 4.2. In Sections 4.3 and 4.4, we provide two additional estimators that have a lower computational cost. All the estimators are provided for the joint predictive distribution (3.4) and the expressions for the predictor and filter are deferred to Section B.1 of the appendix.

**4.1. Term by term variance estimator.** For any  $t \in \mathbb{N}$ , let  $\mathcal{B}_t := \{0, 1\}^{t+1}$ . Denote by  $\mathbf{0}$  the null vector in  $\mathcal{B}_t$  and  $e_s$  the vector with 1 at position  $s$  and 0 elsewhere. Let  $(X_s, X'_s)_{s \in [0:t]}$  be a bivariate



Markov chain in  $(\mathsf{X}^2, \mathcal{X}^{\otimes 2})$  and depending on  $b \in \mathcal{B}_t$  with initial distribution and transition kernels

$$\begin{aligned} \mathcal{M}_0^{b_0}(\mathrm{d}x_0, \mathrm{d}x'_0) &:= \mathrm{M}_0(\mathrm{d}x_0) \{ \mathbb{1}_{b_0=0} \mathrm{M}_0(\mathrm{d}x'_0) + \mathbb{1}_{b_0=1} \delta_{x_0}(\mathrm{d}x'_0) \} , \\ \mathcal{M}_t^{b_t}(x, x'; \mathrm{d}z, \mathrm{d}z') &:= \mathrm{M}_t(x, \mathrm{d}z) \{ \mathbb{1}_{b_t=0} \mathrm{M}_t(x', \mathrm{d}z') + \mathbb{1}_{b_t=1} \delta_z(\mathrm{d}z') \} , \quad \forall t \geq 1 . \end{aligned} \quad (4.1)$$

Define also for any  $b \in \mathcal{B}_t$  the measure  $\mathcal{Q}_{b,t}$  by  $\mathcal{Q}_{b,0}(\mathrm{d}x_0, \mathrm{d}x'_0) = \mathcal{M}_0^{b_0}(\mathrm{d}x_0, \mathrm{d}x'_0)$  and for  $t \geq 1$ :

$$\mathcal{Q}_{b,t}(\mathrm{d}x_{0:t}, \mathrm{d}x'_{0:t}) := \mathcal{M}_0^{b_0}(\mathrm{d}x_0, \mathrm{d}x'_0) \prod_{s=0}^{t-1} g_s^{\otimes 2}(x_s, x'_s) \prod_{s=1}^t \mathcal{M}_s^{b_s}(x_{s-1}, x'_{s-1}; \mathrm{d}x_s, \mathrm{d}x'_s) . \quad (4.2)$$

The measure  $\mathcal{Q}_{b,t}$  is the joint distribution (3.1) of the Feynman-Kac model defined by the initial distribution  $\mathcal{M}_0^{b_0}$ , the transition kernels  $\{\mathcal{M}_s^{b_s}\}_{s \in [1:t]}$  and by the potential functions  $\{g_s^{\otimes 2}\}_{s \in [1:t-1]}$ . Remark that for any  $h \in \mathbb{F}(\mathcal{X})$  such that  $h : x_{0:t} \mapsto h(x_t)$ , we have that  $\mathcal{Q}_{0,t}(h^{\otimes 2}) = \gamma_t(h)^2$  and

$$\mathcal{Q}_{e_s,t}(h^{\otimes 2}) = \gamma_s(\mathbf{1}) \gamma_s(\mathbf{Q}_{s+1:t}[h]^2) . \quad (4.3)$$

A generalization of (4.3) is proved in Proposition 5.1. Consequently, for  $h \in \mathbb{F}_b(\mathcal{X})$   $\mathcal{V}_{\gamma,t}^\infty(h)$  in (3.7) can be rewritten as

$$\mathcal{V}_{\gamma,t}^\infty(h) = \sum_{s=0}^t \{ \mathcal{Q}_{e_s,t}(h_t^{\otimes 2}) - \mathcal{Q}_{0,t}(h_t^{\otimes 2}) \} , \quad (4.4)$$

where  $h_t : x_{0:t} \mapsto h(x_t)$ . Following this observation, for a given  $b$ , an estimator of  $\mathcal{Q}_{b,t}(h_t)$  could be obtained with a single run of a bootstrap particle filter in augmented dimension (i.e. relying on the the bivariate transition  $\mathcal{M}_t^{b_t}$  at time  $t$  and on the weighing of the associated particles with  $g_t^{\otimes 2}$ ). As a direct extension of (3.4), this estimator would be unbiased and as a byproduct, we would get an unbiased estimator of (4.4). However, this procedure is not in line with our initial objective in the sense that we aim at estimating (4.4) with the particles and indices already available.

In order to motivate and introduce our approach, let us consider the static situation where  $t = 0$ ,  $b_0 = 0$  and let  $(h, f) \in \mathbb{F}_b(\mathcal{X})^2$ . Then,

$$\frac{1}{N(N-1)} \sum_{i,j \in [N]^2} \mathbb{1}_{i \neq j} h(\xi_0^i) f(\xi_0^j), \quad \xi_0^{1:N} \stackrel{\text{iid}}{\sim} \mathrm{M}_0 \quad (4.5)$$

is an unbiased and almost sure convergent estimator of  $\mathcal{Q}_{0,0}(h \otimes f) = \mathrm{M}_0^{\otimes 2}(h \otimes f)$  and only relies on i.i.d. samples from  $\mathrm{M}_0$  rather than  $\mathrm{M}_0^{\otimes 2}$ . Note that for  $h = f$ , we thus get an unbiased estimator of  $\gamma_0(h)^2$ . If  $b_0 = 1$ , then  $N^{-1} \sum_{i=1}^N h(\xi_0^i) f(\xi_0^i)$  is an unbiased and consistent estimator of  $\mathcal{Q}_{1,0}(h \otimes f) = \mathrm{M}_0(hf)$ .

Taking advantage of the fact that the particles at time  $t$  are i.i.d. conditionally on  $\mathcal{F}_{t-1}^N$ , we can carry these simple observations to the sequential case in two directions as we now detail. Define for any  $t \in \mathbb{N}^*$  the functional version of the backward weights:

$$\beta_t^N(x, y) := \frac{g_{t-1}(y) m_t(y, x)}{\sum_{\ell=1}^N \omega_{t-1}^\ell m_t(\xi_{t-1}^\ell, x)}, \quad \forall (x, y) \in \mathsf{X}^2 . \quad (4.6)$$

Assume that  $t = 1$  and define for any  $(k_0^1, k_0^2) \in [N]^2$  the following random variables that involve the backward weights

$$\mathcal{E}_0^{\text{BS}}(k_0^1, k_0^2) := \frac{\Omega_0^2}{N(N-1)} \sum_{k_1^{1:2} \in [N]^2} \mathbb{1}_{k_1^1 \neq k_1^2} \beta_1^N(\xi_1^{k_1^1}, \xi_0^{k_1^1}) \beta_1^N(\xi_1^{k_1^2}, \xi_0^{k_1^2}) h(\xi_1^{k_1^1}) f(\xi_1^{k_1^2}), \quad (4.7)$$

$$\mathcal{E}_1^{\text{BS}}(k_0^1, k_0^2) := \frac{\Omega_0^2}{N} \sum_{k_1^{1:2} \in [N]^2} \mathbb{1}_{k_1^1 = k_1^2} \beta_1^N(\xi_1^{k_1^1}, \xi_0^{k_1^1}) \mathcal{W}_0^{k_0^2} h(\xi_1^{k_1^1}) f(\xi_1^{k_1^2}), \quad (4.8)$$

and also the following which involve the ancestors

$$\mathcal{E}_0^{\text{GT}}(k_0^1, k_0^2) := \frac{\Omega_0^2}{N(N-1)} \sum_{k_1^{1:2} \in [N]^2} \mathbb{1}_{A_0^{k_1^1} = k_0^1, A_0^{k_1^2} = k_0^2, k_1^1 \neq k_1^2} h(\xi_1^{k_1^1}) f(\xi_1^{k_1^2}), \quad (4.9)$$

$$\mathcal{E}_1^{\text{GT}}(k_0^1, k_0^2) := \frac{\Omega_0^2}{N} \sum_{k_1^{1:2} \in [N]^2} \mathbb{1}_{k_0^1 = A_0^{k_1^1}, k_1^1 = k_1^2} \mathcal{W}_0^{k_0^2} h(\xi_1^{k_1^1}) f(\xi_1^{k_1^2}), \quad (4.10)$$

Consider also Lemma 4.1 which states a crucial identity involving the backward weights.

**Lemma 4.1.** *For all  $t \in \mathbb{N}^*$ ,  $(x, y) \in \mathcal{X}^2$ ,*

$$\beta_t^N(x, y) \phi_{t-1}^N \mathbf{M}_t(\mathrm{d}x) = \frac{g_{t-1}(y)}{\Omega_{t-1}} \mathbf{M}_t(y, \mathrm{d}x). \quad (4.11)$$

and for any  $(k_{t-1}^1, k_t^1) \in [N]^2$  and  $h \in \mathbb{F}(\mathcal{X})$ ,

$$\mathbb{E}[\beta_t^N(\xi_t^{k_t}, \xi_{t-1}^{k_{t-1}}) h(\xi_t^{k_t}) | \mathcal{F}_{t-1}^N] = \mathbb{E}[\mathbb{1}_{k_{t-1} = A_{t-1}^{k_t}} h(\xi_t^{k_t}) | \mathcal{F}_{t-1}^N] = \mathcal{W}_{t-1}^{k_{t-1}} \mathbf{M}_t[h](\xi_{t-1}^{k_{t-1}}). \quad (4.12)$$

Applying Lemma 4.1 to (4.7) and (4.9) and using that given  $\mathcal{F}_0^N$  the particles at  $t = 1$  are i.i.d., we get

$$\mathbb{E}[\mathcal{E}_0^{\text{BS}}(k_0^1, k_0^2) | \mathcal{F}_0^N] = \mathbb{E}[\mathcal{E}_0^{\text{GT}}(k_0^1, k_0^2) | \mathcal{F}_0^N] = g_0^{\otimes 2}(\xi_0^{k_0^1}, \xi_0^{k_0^2}) \mathcal{M}_1^0[h \otimes f](\xi_0^{k_0^1}, \xi_0^{k_0^2}),$$

and

$$\begin{aligned} \mathbb{E}\left[\sum_{k_0^{1:2} \in [N]^2} \mathbb{1}_{k_0^1 \neq k_0^2} \mathcal{E}_0^{\text{BS}}(k_0^1, k_0^2)\right] &= \mathbb{E}\left[\sum_{k_0^{1:2} \in [N]^2} \mathbb{1}_{k_0^1 \neq k_0^2} \mathbb{E}[\mathcal{E}_0^{\text{BS}}(k_0^1, k_0^2) | \mathcal{F}_0^N]\right] \\ &= \mathbb{E}\left[\sum_{k_0^{1:2} \in [N]^2} \mathbb{1}_{k_0^1 \neq k_0^2} \mathbb{E}[\mathcal{E}_0^{\text{GT}}(k_0^1, k_0^2) | \mathcal{F}_0^N]\right] \\ &= N(N-1) \mathcal{Q}_{0,1}(h \otimes f). \end{aligned}$$

Similarly,

$$\mathbb{E}[\mathcal{E}_1^{\text{BS}}(k_0^1, k_0^2) | \mathcal{F}_0^N] = \mathbb{E}[\mathcal{E}_1^{\text{GT}}(k_0^1, k_0^2) | \mathcal{F}_0^N] = g_0^{\otimes 2}(\xi_0^{k_0^1}, \xi_0^{k_0^2}) \mathcal{M}_1^1[h \otimes f](\xi_0^{k_0^1}, \xi_0^{k_0^2});$$

and

$$\mathbb{E}\left[\sum_{k_0^{1:2} \in [N]^2} \mathbb{1}_{k_0^1 \neq k_0^2} \mathcal{E}_1^{\text{BS}}(k_0^1, k_0^2)\right] = N(N-1) \mathcal{Q}_{e,1}(h \otimes f).$$

Therefore, it is possible to derive unbiased estimators of  $\mathcal{Q}_{0,1}(h \otimes f)$  and  $\mathcal{Q}_{e,1}(h \otimes f)$  (but also  $\mathcal{Q}_{e_0,1}(h \otimes f)$  and  $\mathcal{Q}_{(1,1),1}(h \otimes f)$ ) with a single run of the particle filter in two different ways: either by using the backward weights and (4.7)-(4.8), or by using directly the ancestry of the particles and

(4.9)-(4.10). The asymptotic variance estimators proposed in [4, 27, 32, 18] are all based on the latter solution, while in this paper we instead focus on estimators based on the backward weights.

We now generalize the derivations performed in the case  $t = 1$ . Denote by  $\Lambda_{1,t}$  and  $\Lambda_{2,t}$  the discrete measures conditioned on  $\mathcal{F}_t^N$  and defined by

$$\Lambda_{1,t}(\mathrm{d}k_{0:t}) := N^{-1} \prod_{s=1}^t \beta_s(k_s, \mathrm{d}k_{s-1}), \quad (4.13)$$

$$\Lambda_{2,t}(k_{0:t}^1; \mathrm{d}k_{0:t}^2) := N^{-1} \prod_{s=1}^t \left\{ \mathbb{1}_{k_s^2 = k_s^1} \mathcal{W}_{s-1}^{k_s^2-1} + \mathbb{1}_{k_s^2 \neq k_s^1} \beta_s(k_s^2, \mathrm{d}k_{s-1}^2) \right\}. \quad (4.14)$$

Specific choices of kernels  $\{\beta_s\}_{s=1}^t$  are, for all  $(k, \ell) \in [N]^2$ ,

$$\beta_s^{\text{GT}}(k, \ell) := \mathbb{1}_{\ell = A_{s-1}^k}, \quad \beta_s^{\text{BS}}(k, \ell) := \frac{\omega_{s-1}^\ell m_s(\xi_{s-1}^\ell, \xi_s^k)}{\sum_{j=1}^N \omega_{s-1}^j m_s(\xi_{s-1}^j, \xi_s^k)} = \beta_s^N(\xi_s^k, \xi_{s-1}^\ell),$$

where GT stands for *genealogy tracing* and BS for *backward sampling*. When the conditional distribution given  $\mathcal{F}_t^N$  is  $\Lambda_1^{\text{BS}} \otimes \Lambda_2^{\text{BS}}$  (resp.  $\Lambda_1^{\text{GT}} \otimes \Lambda_2^{\text{GT}}$ ) we write  $\mathbb{E}_{\text{BS}}[\cdot | \mathcal{F}_t^N]$  (resp.  $\mathbb{E}_{\text{GT}}[\cdot | \mathcal{F}_t^N]$ ). Define also for any  $b \in \mathcal{B}_t$  the coalescence function:

$$\mathbf{I}_{b,s} : ([N]^{s+1})^2 \ni (k_{0:s}^1, k_{0:s}^2) \mapsto \prod_{\ell=0}^s \{ \mathbb{1}_{k_\ell^1 = k_\ell^2} \mathbb{1}_{b_\ell=1} + \mathbb{1}_{k_\ell^1 \neq k_\ell^2} \mathbb{1}_{b_\ell=0} \}, \quad \forall s \in [0 : t], \quad (4.15)$$

for any  $h \in \mathbb{F}(\mathcal{X}^{\otimes 2(t+1)})$  the random variable

$$\mathcal{Q}_{b,t}^{N,\text{BS}}(h) := \prod_{s=0}^t N^{b_s} \left( \frac{N}{N-1} \right)^{1-b_s} \gamma_t^N(\mathbf{1})^2 \mathbb{E}_{\text{BS}}[\mathbf{I}_{b,t}(K_{0:t}^1, K_{0:t}^2) h(\xi_{0:t}^{K_{0:t}^1}, \xi_{0:t}^{K_{0:t}^2}) | \mathcal{F}_t^N], \quad (4.16)$$

and denote by  $\mathcal{Q}_{b,t}^{N,\text{GT}}$  the counterpart where the expectation in the r.h.s. is  $\mathbb{E}_{\text{GT}}$ .

**Remark 4.2.** By (4.15), the random variable  $\mathcal{Q}_{b,t}^{N,\text{BS}}(h)$  remains defined for any  $b \in \mathcal{B}_r$  with  $r > t$  and  $\mathcal{Q}_{b,t}^{N,\text{BS}}(h) = \mathcal{Q}_{b_{0:t},t}^{N,\text{BS}}(h)$  where  $b_{0:t}$  is the truncation of  $b$  to the  $t+1$  first terms.

Finally, define for any  $h \in \mathbb{F}(\mathcal{X})$ ,

$$\bar{\mathcal{V}}_{\gamma,t}^{N,\text{BS}}(h) := \sum_{s=0}^t \{ \mathcal{Q}_{e_s,t}^{N,\text{BS}}(h^{\otimes 2}) - \mathcal{Q}_{0,t}^{N,\text{BS}}(h^{\otimes 2}) \}. \quad (4.17)$$

**Proposition 4.3.** *Let  $t \in \mathbb{N}$ . For any  $b \in \mathcal{B}_t$  and any  $h \in \mathbb{F}(\mathcal{X}^{\otimes 2(t+1)})$ ,*

- (i)  $\mathbb{E}[\mathcal{Q}_{b,t}^{N,\text{BS}}(h) | \mathcal{F}_{t-1}^N] = \mathcal{Q}_{b,t-1}^{N,\text{BS}}(g_{t-1}^{\otimes 2} \mathcal{M}_t^{b_t}[h])$  for all  $t \in \mathbb{N}^*$ .
- (ii)  $\mathcal{Q}_{b,t}^{N,\text{BS}}(h)$  is an unbiased estimator of  $\mathcal{Q}_{b,t}(h)$ .
- (iii) If  $h \in \mathbb{F}(\mathcal{X})$ ,  $\bar{\mathcal{V}}_{\gamma,t}^{N,\text{BS}}(h)$  is an unbiased estimator of  $\mathcal{V}_{\gamma,t}^\infty(h)$ .

The proof is provided in Section A.2 of the appendix. First, (ii) is a generalization of [27, Lemma 2] which states that  $\mathcal{Q}_{b,t}^{N,\text{GT}}(h)$  is also an unbiased estimator  $\mathcal{Q}_{b,t}(h)$ . Its proof (see [27, Supplementary]) is based on a doubly conditional SMC argument [2] and while this scheme can be replicated to our estimator based on backward weights, we rather propose an alternative and elementary proof that also extends straightforwardly to GT and which is based on our previous discussion. From (4.4), (iii) is a direct consequence of (ii) and provides an estimator of (4.4) based on a single particle run.

Theorem 4.4 deals with the convergence of  $\mathcal{Q}_{b,t}^{N,\text{BS}}(h)$  for bounded  $h$ . The convergence in  $\mathbf{L}_2$  is stated under (A2 : 4) which are standard and the  $1/\sqrt{N}$  convergence rate is obtained under the additional assumption (A5). The equivalent result for  $\mathcal{Q}_{b,t}^{N,\text{GT}}$  is stated in [27] and is proved under (A2) alone. From a technical point of view, this is possible because the use of indicators instead of backward weights allows for cancellations that simplify the analysis significantly. Assumption (A4) enables us to show that the additional terms that come with the use of backward weights go to zero.

(A3) For all  $t > 0$  and  $(x, x') \in \mathbf{X}^2$ ,  $m_t(x', x) > 0$ .

(A4) There exists  $\sigma_+ > 0$  such that for all  $t \geq 1$ ,  $\sup_{x, x' \in \mathbf{X}} m_t(x', x) \leq \sigma_+$ .

(A5) There exists  $0 < \sigma_- < \sigma_+$  such that for all  $t \geq 1$ ,  $\inf_{x, x' \in \mathbf{X}} m_t(x', x) \geq \sigma_-$ .

Assumption (A5) is a strong assumption that is typically verified in models where the state space  $\mathbf{X}$  is compact. This assumption, together with (A4), are now classic and have been widely used to obtain quantitative bounds in the SMC literature [19, 15, 28].

**Theorem 4.4.** Assume that (A2 : 4) hold. For any  $t \in \mathbb{N}$ ,  $b \in \mathcal{B}_t$  and  $h \in \mathbb{F}_b(\mathcal{X}^{\otimes 2(t+1)})$ ,

$$\lim_{N \rightarrow \infty} \|\mathcal{Q}_{b,t}^{N,\text{BS}}(h) - \mathcal{Q}_{b,t}(h)\|_2 = 0. \quad (4.18)$$

In addition, if (A5) holds the convergence rate is  $\mathcal{O}(1/\sqrt{N})$ .

The proof can be found in Section A.4 of the appendix. As a straightforward consequence, the term by term estimator (4.17) of the asymptotic variance is weakly consistent. It remains to detail how it can be computed. The next section is devoted to the exact computation of the estimators  $\mathcal{Q}_{0,t}^{N,\text{BS}}(h)$  and  $\mathcal{Q}_{e_s,t}^{N,\text{BS}}(h)$  that appear in its expression.

**4.2. Computation for  $b = 0$  and  $b = e_s$ .** We now derive practical expressions of  $\mathcal{Q}_{0,t}^{N,\text{BS}}(h)$  and  $\mathcal{Q}_{e_s,t}^{N,\text{BS}}(h)$  in the practical case where  $h : x_{0:t}, x'_{0:t} \mapsto h(x_t, x'_t) \in \mathbb{F}(\mathcal{X}^{\otimes 2(t+1)})$ . Define, for any  $b \in \mathcal{B}_t$  and any  $t \geq 0$ ,

$$\mathcal{T}_t^b(K_t^1, K_t^2) := \mathbb{E}_{\text{BS}}[\mathbb{I}_{b,t}(K_{0:t}^1, K_{0:t}^2) | \mathcal{F}_t^N, K_t^1, K_t^2]. \quad (4.19)$$

Then, by the tower property,  $\mathcal{Q}_{b,t}^{N,\text{BS}}(h)$  in (4.16) can be rewritten as

$$\mathcal{Q}_{b,t}^{N,\text{BS}}(h) = \prod_{s=0}^t N^{b_s} \left( \frac{N}{N-1} \right)^{1-b_s} \frac{\gamma_t^N(\mathbf{1})^2}{N^2} \sum_{k, \ell \in [N]^2} \mathcal{T}_t^b(k, \ell) h(\xi_t^k, \xi_t^\ell). \quad (4.20)$$

Next, define for any  $t \in \mathbb{N}$  and  $(k, \ell) \in [N]^2$ ,

$$S_t(k, \ell) := \sum_{s=0}^t \mathcal{T}_s^{e_s}(k, \ell). \quad (4.21)$$

Plugging (4.20) in (4.17),  $\bar{\mathcal{V}}_{\gamma,t}^{N,\text{BS}}(h)$  can be rewritten as

$$\bar{\mathcal{V}}_{\gamma,t}^{N,\text{BS}}(h) = \frac{N^{t-1} \gamma_t^N(\mathbf{1})^2}{(N-1)^t} \sum_{k, \ell \in [N]^2} \left\{ S_t(k, \ell) - \frac{t+1}{N-1} \mathcal{T}_t^0(k, \ell) \right\} h(\xi_t^k) h(\xi_t^\ell). \quad (4.22)$$

The sequential computation of  $\bar{\mathcal{V}}_{\gamma,t}^{N,\text{BS}}(h)$  relies on that of  $S_t(k, \ell)$  in (4.21), and so on that of  $\mathcal{T}_t^{e_s}(k, \ell)$  and  $\mathcal{T}_t^0(k, \ell)$ . By the tower property, we obtain the following recursions for  $\mathcal{T}_t^b$ :

$$\begin{cases} \mathcal{T}_0^b(k, \ell) = \mathbb{1}_{k \neq \ell, b_0=0} + \mathbb{1}_{k=\ell, b_0=1} , \\ \mathcal{T}_t^b(k, \ell) = \mathbb{1}_{k \neq \ell} \sum_{i,j \in [N]^2} \beta_t^{\text{BS}}(k, i) \beta_t^{\text{BS}}(\ell, j) \mathcal{T}_{t-1}^b(i, j) & \text{if } b_t = 0 , \\ \mathcal{T}_t^b(k, \ell) = \mathbb{1}_{k=\ell} \sum_{i,j \in [N]^2} \beta_t^{\text{BS}}(k, i) \mathcal{W}_{t-1}^j \mathcal{T}_{t-1}^b(i, j) & \text{if } b_t = 1 , \end{cases} \quad (4.23)$$

for all  $(k, \ell) \in [N]^2$  and  $t \in \mathbb{N}^*$ . In particular, if  $b = 0$ ,

$$\mathcal{T}_t^0(k, \ell) = \mathbb{1}_{k \neq \ell} \sum_{i,j \in [N]^2} \beta_t^{\text{BS}}(k, i) \beta_t^{\text{BS}}(\ell, j) \mathcal{T}_{t-1}^0(i, j) , \quad (4.24)$$

and if  $b = e_s$ ,

$$\mathcal{T}_t^{e_s}(k, \ell) = \begin{cases} \mathbb{1}_{k \neq \ell} \sum_{i,j \in [N]^2} \beta_t^{\text{BS}}(k, i) \beta_t^{\text{BS}}(\ell, j) \mathcal{T}_{t-1}^{e_s}(i, j) & t > s , \\ \mathbb{1}_{k=\ell} \sum_{i,j \in [N]^2} \beta_t^{\text{BS}}(k, i) \mathcal{W}_{t-1}^j \mathcal{T}_{t-1}^0(i, j) & t = s , \\ \mathcal{T}_t^0(k, \ell) & t < s . \end{cases} \quad (4.25)$$

Next, combining (4.21)-(4.25) we obtain the online update of  $S_t$ :

$$S_t(k, \ell) = \mathcal{T}_t^{e_t}(k, \ell) + \mathbb{1}_{k \neq \ell} \sum_{i,j \in [N]^2} \beta_t^{\text{BS}}(k, i) \beta_t^{\text{BS}}(\ell, j) S_{t-1}(i, j) , \quad (4.26)$$

for any  $(k, \ell) \in [N]^2$  and  $t \in \mathbb{N}$ . We have shown that despite the sum over  $s$  that appears in (4.17) we are still able to update (4.22) at a computational cost independent of the time horizon  $t$  by propagating  $S_t$  and  $\mathcal{T}_t^0$ . Note that the algorithm provided in [27, Supplementary] for the computation of  $\bar{\mathcal{V}}_{\gamma,t}^{N,\text{GT}}$  relies on that of each individual term that appears in the GT version of the sum (4.17) and is not tractable from an online point of view. We thus derive a similar online scheme for the GT estimator in Section B.2 of the appendix.

The computation of the estimates  $\mathcal{Q}_{b,t}^{N,\text{BS}}(h)$  and  $\bar{\mathcal{V}}_{\gamma,t}^{N,\text{BS}}(h)$  can benefit from parallelization by implementing the updates (4.24)-(4.25) with matrix operations. They then become

$$\begin{cases} \mathcal{T}_t^b = \beta_t^{\text{BS}} \mathcal{T}_{t-1}^b \beta_t^{\text{BS}\top} - \text{Diag}(\beta_t^{\text{BS}} \mathcal{T}_{t-1}^b \beta_t^{\text{BS}\top}) & \text{if } b_t = 0 , \\ \mathcal{T}_t^b = \text{Diag}(\beta_t^{\text{BS}} \mathcal{T}_{t-1}^b \mathcal{W}_{t-1}^{1:N}) & \text{if } b_t = 1 . \end{cases}$$

**4.3. Variance estimators with reduced computational cost.** In this section we derive a second estimator that relies only on the update of  $\mathcal{T}_t^0$ . Let  $h \in \mathbb{F}_b(\mathcal{X})$ . By (3.6),  $\sqrt{N}(\gamma_t^N(h) - \gamma_t(h))$  converges in distribution; moreover,  $N(\gamma_t^N(h) - \gamma_t(h))^2$  is uniformly integrable, using for instance a Hoeffding type inequality (see [16]). Hence  $N\mathbb{E}[(\gamma_t^N(h) - \gamma_t(h))^2]$  converges to the asymptotic variance  $\mathcal{V}_{\gamma,t}^\infty(h)$ . On the other hand, using the lack of bias of  $\gamma_t^N(h)$ ,

$$N\mathbb{E}[(\gamma_t^N(h) - \gamma_t(h))^2] = N(\mathbb{E}[\gamma_t^N(h)^2] - \gamma_t(h)^2) = N(\mathbb{E}[\gamma_t^N(h)^2] - \mathcal{Q}_{0,t}(h^{\otimes 2})) .$$

A natural estimator of this quantity is obtained by replacing both terms by their unbiased estimators  $\gamma_t^N(h)^2$  and  $\mathcal{Q}_{0,t}^{N,\text{BS}}(h^{\otimes 2})$

$$\begin{aligned} \mathcal{V}_{\gamma,t}^{N,\text{BS}}(h) &:= N(\gamma_t^N(h)^2 - \mathcal{Q}_{0,t}^{N,\text{BS}}(h^{\otimes 2})) \\ &= N\gamma_t^N(\mathbf{1})^2 \left( \eta_t^N(h)^2 - \frac{N^{t-1}}{(N-1)^{t+1}} \sum_{i,j \in [N]^2} \mathcal{T}_t^0(i, j) h(\xi_t^i) h(\xi_t^j) \right) . \end{aligned} \quad (4.27)$$

The convergence of this second estimator stated in Theorem 4.5 stems from the following identity which also appears in [27, 18] and dates back to [8]:

$$\begin{aligned} & \sum_{b \in \mathcal{B}_t} \left\{ \prod_{s=0}^t \frac{1}{N^{b_s}} \left( \frac{N-1}{N} \right)^{1-b_s} \right\} \mathcal{Q}_{b,t}^{N,BS}(h_t^{\otimes 2}) \\ &= \gamma_t^N(\mathbf{1})^2 \mathbb{E}_{BS} \left[ \sum_{b \in \mathcal{B}_t} \mathbf{I}_{b,t}(K_{0:t}^1, K_{0:t}^2) h(\xi_t^{K_t^1}) h(\xi_t^{K_t^2}) \middle| \mathcal{F}_t^N \right] = \gamma_t^N(\mathbf{1})^2 \eta_t^N(h)^2 = \gamma_t^N(h)^2. \end{aligned} \quad (4.28)$$

**Theorem 4.5.** *Let (A2 : 4) hold. For any  $h \in \mathbb{F}_b(\mathcal{X})$ ,  $\mathcal{V}_{\gamma,t}^{N,BS}(h)$  converges in probability to  $\mathcal{V}_{\gamma,t}^\infty(h)$ .*

The proof is in Section A.5 of the appendix. The main advantage of (4.27) w.r.t. (4.22) is the computational cost. Indeed, remark that (4.27) only relies on the sequential update of  $\mathcal{T}_t^0$ , contrary to (4.22) which also relies on that of  $\mathcal{T}_t^{e_s}$ . Consequently, the computational time of (4.27) is roughly twice lower. In addition, experiments show that the difference in performance is negligible so (4.27) is to be preferred in practice.

*Remark 4.6.* This alternative estimator does not invalidate the relevance of (4.17). Indeed, remember that (4.17) is an unbiased estimator. Moreover, the asymptotic variance estimator of the FFBS algorithm that we provide in Section 5.2 is a term by term estimator that can be updated online in a way similar to (4.17).

**4.4. A PaRIS variance estimator.** Let us discuss how the computational cost of (4.27) and (4.22) can be further reduced *à la PaRIS* [33, 21]. In [33], the forward only implementation of the FFBS algorithm is sped up by replacing the backward statistics by a conditionally unbiased estimator obtained by sampling particle indices according to the backward probabilities  $\beta_t^{BS}$  through rejection sampling. We therefore apply the same idea here by letting  $\tilde{\mathcal{T}}_0^0 := \mathcal{T}_0^0$  and replacing  $\mathcal{T}_t^b$  with

$$\begin{aligned} \tilde{\mathcal{T}}_t^b(k, \ell) &:= \frac{\mathbb{1}_{k \neq \ell}}{M} \sum_{i=1}^M \tilde{\mathcal{T}}_{t-1}^b(J_{k,t-1}^i, J_{\ell,t-1}^i) \quad \text{if } b_t = 0, \\ \tilde{\mathcal{T}}_t^b(k, \ell) &:= \frac{\mathbb{1}_{k=\ell}}{M} \sum_{i=1}^M \sum_{j=1}^N \mathcal{W}_{t-1}^j \tilde{\mathcal{T}}_{t-1}^b(J_{k,t-1}^i, j) \quad \text{if } b_t = 1, \end{aligned}$$

where for any  $k \in [N]$ ,  $J_{k,t-1}^{1:M}$  are i.i.d. samples according to  $\beta_t^{BS}(k, \cdot)$ . For  $h \in \mathbb{F}(\mathcal{X}^{\otimes 2})$ , the *PaRIS* estimator of  $\mathcal{Q}_{b,t}(h)$  is, for any  $b \in \mathcal{B}_t$

$$\tilde{\mathcal{Q}}_{b,t}^{N,M}(h) = \left\{ \prod_{s=0}^t N^{b_s} \left( \frac{N}{N-1} \right)^{1-b_s} \right\} \frac{\gamma_t^N(\mathbf{1})^2}{N^2} \sum_{i,j \in [N]^2} \tilde{\mathcal{T}}_t^b(i, j) h(\xi_t^i, \xi_t^j), \quad (4.29)$$

and the *PaRIS* variance estimators are

$$\bar{\mathcal{V}}_{\gamma,t}^{N,M}(h) = \sum_{s=0}^t \{ \tilde{\mathcal{Q}}_{e_s,t}^{N,M}(h^{\otimes 2}) - \tilde{\mathcal{Q}}_{\mathbf{0},t}^{N,M}(h^{\otimes 2}) \}, \quad (4.30)$$

$$\mathcal{V}_{\gamma,t}^{N,M}(h) = N(\gamma_t^N(h)^2 - \tilde{\mathcal{Q}}_{\mathbf{0},t}^{N,M}(h^{\otimes 2})), \quad (4.31)$$

where  $M > 1$  refers to the number of sampled indices. The computation of (4.27) and (4.31) is summarized in Algorithm 1.

---

**Algorithm 1** Variance estimators (4.27) and (4.31) associated to  $\gamma_t^N(h)$ 


---

**Require:**  $M, \omega_t^{1:N}, \xi_t^{1:N}, \xi_{t+1}^{1:N}, \mathcal{T}_t^0$  and  $\gamma_t^N(1)$

  Compute  $\beta_{t+1}^{\text{BS}}$

**if** *PaRIS* **then**

**for**  $k \in [1 : N]$  **do**

      Sample  $J_{k,t}^{1:M} \stackrel{\text{iid}}{\sim} \beta_{t+1}^{\text{BS}}(k, \cdot)$

**end for**

**for**  $(k, \ell) \in [1 : N]^2$  **do**

      Set  $\mathcal{T}_{t+1}^0(k, \ell) = \mathbb{1}_{k \neq \ell} \sum_{i=1}^M \mathcal{T}_t^0(J_{k,t}^i, J_{\ell,t}^i) / M$

**end for**

**else**

    Compute  $\overline{\mathcal{T}}_{t+1}^0 = \beta_{t+1}^{\text{BS}} \mathcal{T}_t^0 \beta_{t+1}^{\text{BS} \prime}$ .

    Set  $\mathcal{T}_{t+1}^0 = \overline{\mathcal{T}}_{t+1}^0 - \text{Diag}(\overline{\mathcal{T}}_{t+1}^0)$ .

**end if**

  Compute  $\mathcal{Q} = \mathcal{T}_{t+1}^0 \odot [h(\xi_{t+1}^{1:N})h(\xi_{t+1}^{1:N})^\top]$ .  $\triangleright h$  is applied elementwise

**return**  $N\gamma_t^N(1)^2 \{ \eta_t^N(h)^2 - N^{t-2} \sum_{i,j \in [N]^2} \mathcal{Q}_{i,j} / (N-1)^{t+1} \}, \quad \mathcal{T}_{t+1}^0$ .

---

We are able to reduce the time complexity of computing  $\mathcal{T}_t^b$  to  $\mathcal{O}(MN^2)$ . The key feature of the *PaRIS* approach is that  $M$  does not necessarily need to be large (see [33, Section 3.1] for a discussion on this matter). We impose  $M > 1$  because then in the case  $b = 0$ , which is the case we are the most interested in, the support of  $\mathcal{Q}_{0,t}^{N,\text{BS}}(h)$  is made of  $N^2 M^{t+1}$  terms whereas when  $M = 1$  it is only  $N^2$ . We show empirically in our experiments that setting  $M = 3$  is sufficient to provide good results for the asymptotic variance estimation.

While it is not needed to obtain a  $\mathcal{O}(MN^2)$  time complexity, the indices  $J_{k,t-1}^{1:M}$  can be sampled using an accept-reject procedure with the weights  $\mathcal{W}_t^{1:N}$  as proposals if the transition densities  $m_t$  are upper bounded. This approach does not require the computation of the normalizing constant of the backward weights (4.6). The computational time is then random but if the transition kernels are strongly mixing it can be provably further reduced [15]. Theorem 4.7 is concerned with the convergence of  $\tilde{\mathcal{Q}}_{b,t}^{N,M}(h)$  for any bounded  $h$  and for any fixed  $M > 1$ . Its proof bears some similarity with that of Theorem 4.4 with the exception that the additional sampling introduces non trivial terms that need to be handled carefully. As a straightforward consequence, we obtain the convergence in probability of  $\bar{\mathcal{V}}_{\gamma,t}^{N,M}(h)$  for any  $h \in \mathbb{F}_b(\mathcal{X})$ . The weak consistency of (4.31) in Theorem 4.8 is however less straightforward than that of Theorem 4.5 and relies on the insight that the identity (4.28) still holds when  $\mathcal{Q}_{b,t}^{N,\text{BS}}$  are replaced with their *PaRIS* versions. The proofs are provided respectively in Section A.6 and A.7 of the appendix.

**Theorem 4.7.** Assume that (A2 : 4) hold. For any  $t \in \mathbb{N}$ ,  $b \in \mathcal{B}_t$ ,  $M > 1$  and  $h \in \mathbb{F}_b(\mathcal{X}^{\otimes 2})$ ,

$$\lim_{N \rightarrow \infty} \|\tilde{\mathcal{Q}}_{b,t}^{N,M}(h) - \mathcal{Q}_{b,t}(h)\|_2 = 0. \quad (4.32)$$

In addition, if (A5) holds the convergence rate is  $\mathcal{O}(1/\sqrt{N})$ .

**Theorem 4.8.** Let (A2 : 4) hold. For all  $t \in \mathbb{N}$ ,  $M > 1$  and  $h \in \mathbb{F}_b(\mathcal{X})$ ,  $\mathcal{V}_{\gamma,t}^{N,M}(h)$  converges in probability to  $\mathcal{V}_{\gamma,t}^\infty(h)$  when  $N$  goes to infinity.

## 5. APPLICATION TO THE FFBS

In this section, we derive an estimator for the asymptotic variance of the *Forward Filtering Backward Smoothing* algorithm. We start by giving a short presentation of the FFBS algorithm and we next derive an estimator of the asymptotic variance for additive functionals.

**5.1. FFBS algorithm.** The FFBS algorithm aims at solving the well known degeneracy problem associated with the particle filter of Section 3.2 when it is used for approximating smoothing distributions. It relies on the following backward decomposition of the joint smoothing distribution:

$$\phi_{0:t|t}(h) = \int h(x_{0:t}) \phi_t(dx_t) \mathbf{T}_t(x_t, dx_{0:t-1}), \quad (5.1)$$

where  $\mathbf{T}_t$  is the backward transition kernel from  $(\mathbf{X}, \mathcal{X})$  to  $(\mathbf{X}^t, \mathcal{X}^{\otimes t})$ :

$$\mathbf{T}_t := \begin{cases} \mathbf{B}_{\phi_{t-1}} \otimes \cdots \otimes \mathbf{B}_{\phi_0} & \text{if } t > 0, \\ \text{Id} & \text{else,} \end{cases}$$

and  $\mathbf{B}_{\phi_s}$  is the backward kernel defined by

$$\mathbf{B}_{\phi_s}(x_{s+1}, A) := \frac{\int m_{s+1}(x_s, x_{s+1}) \mathbb{1}_A(x_s) \phi_s(dx_s)}{\phi_s(m_{s+1}(\cdot, x_{s+1}))}, \quad \forall A \in \mathcal{X}, \forall x_{s+1} \in \mathbf{X}.$$

Denote by  $\mathbf{T}_t^N$  the particle approximation of  $\mathbf{T}_t$  where each backward kernel  $\mathbf{B}_{\phi_s}$  is replaced by plugging in the particle approximation of the filter. This yields for any  $A \in \mathcal{X}$  and  $x_{s+1} \in \mathbf{X}$ ,

$$\mathbf{B}_{\phi_s}^N(x_{s+1}, A) := \frac{\int m_{s+1}(x_s, x_{s+1}) \mathbb{1}_A(x_s) \phi_s^N(dx_s)}{\phi_s^N(m_{s+1}(\cdot, x_{s+1}))} = \sum_{i=1}^N \frac{\omega_s^i m_{s+1}(\xi_s^i, x_{s+1})}{\sum_{j=1}^N \omega_s^j m_{s+1}(\xi_s^j, x_{s+1})} \mathbb{1}_A(\xi_s^i).$$

Plugging this approximation and that of the filtering distribution filter in (5.1) yields

$$\phi_{0:t|t}^{N, \text{FFBS}}(h) := \sum_{i_0=1}^N \cdots \sum_{i_t=1}^N \tilde{\Lambda}_t(i_{0:t}) h(\xi_{i_0}^{i_0}, \dots, \xi_{i_t}^{i_t}), \quad (5.2)$$

where  $\tilde{\Lambda}_t(i_{0:t}) := \mathcal{W}_t^{i_t} \prod_{s=1}^t \beta_t^{\text{BS}}(i_s, i_{s-1})$ . In the following, we write  $\phi_{0:t|t}^N$  for  $\phi_{0:t|t}^{N, \text{FFBS}}$  and if  $h$  is such that  $h : x_{0:t} \mapsto h(x_{s:\ell})$  with  $0 \leq s \leq \ell \leq t$ , we will instead write  $\phi_{s:\ell|t}^N(h)$ .

The theoretical properties of the FFBS are well understood in both the asymptotic regimes of  $N$  and  $t$  [15, 10, 11, 19, 33, 16]. In particular, a Central Limit Theorem with an explicit expression of the asymptotic variance is established for any  $h \in \mathbb{F}_b(\mathcal{X}^{\otimes t+1})$  under (A2) in [15, Theorem 8],

$$\sqrt{N}(\phi_{0:t|t}^N(h) - \phi_{0:t|t}(h)) \implies \mathcal{N}(0, \mathcal{V}_{0:t|t}^{\text{FFBS}}(h)), \quad (5.3)$$

where

$$\mathcal{V}_{0:t|t}^{\text{FFBS}}(h) := \sum_{s=0}^t \frac{\eta_s(\mathbf{G}_{s,t}[g_t\{h - \phi_{0:t|t}(h)\}]^2)}{\eta_s(\mathbf{Q}_{s+1,t}[g_t])^2} \quad (5.4)$$

and  $\mathbf{G}_{s,t}$  is the kernel that integrates  $h$  forward and backward starting from  $x_s$ , i.e.

$$\mathbf{G}_{s,t}[h](x_s) := \mathbf{T}_s[\mathbf{Q}_{s+1,t}[h]](x_s) = \int h(x_{0:t}) \mathbf{T}_s(x_s, dx_{0:s-1}) \mathbf{Q}_{s+1,t}(x_s, dx_{s+1:t}),$$

for any  $s \in [0 : t]$  and  $x_s \in \mathbf{X}$ .

Unlike the asymptotic variance of filtering algorithms, no estimator of (5.4) exists in the literature, even though the FFBS and its variants are of significant importance in marginal smoothing and parameter estimation in HMMs [24]. For a general  $h$ , (5.2) has  $\mathcal{O}(N^{t+1})$  complexity and is thus



impractical for any problem of interest, especially in an online setting where the observations are processed sequentially. It is therefore not conceivable to be able to devise an estimator that computes the asymptotic variance at a lower computational cost, let alone an estimator that tracks the asymptotic variance online. For this reason, we restrict ourselves to additive functionals  $h$  of the form:

$$h_{0:t}(x_{0:t}) = \sum_{s=0}^{t-1} \tilde{h}_s(x_s, x_{s+1}), \quad (5.5)$$

for which the FFBS can be computed online with a  $\mathcal{O}(N^2)$  time complexity per time step. For  $0 \leq s < r \leq t$ , we write  $\tilde{h}_{s:r}(x_{s:r}) = \sum_{\ell=s}^{r-1} \tilde{h}_\ell(x_\ell, x_{\ell+1})$ . Expectations of functionals of the form (5.5) include marginal smoothing, pairwise marginal smoothing and the  $E$ -step of the Expectation Maximization algorithm.

We show in the next section that it is in fact possible to estimate (5.4) online for such functionals. Before that, let us first recall why the FFBS can be indeed computed online in this case. For more details on the forward only implementation of the FFBS and its variants we refer the reader to [16, 33]. For any  $t > 0$  and any additive functional  $h_{0:t}$ ,

$$\begin{aligned} \mathbf{T}_t[h_{0:t}](x_t) &= \int \left\{ \tilde{h}_{0:t-1}(x_{0:t-1}) + \tilde{h}_t(x_{t-1}, x_t) \right\} \mathbf{T}_t(x_t, dx_{0:t-1}) \\ &= \int \left\{ \tilde{h}_{0:t-1}(x_{0:t-1}) + \tilde{h}_{t-1}(x_{t-1}, x_t) \right\} \mathbf{B}_{\phi_{t-1}}(x_t, dx_{t-1}) \mathbf{T}_{t-1}(x_{t-1}, dx_{0:t-2}) \\ &= \mathbf{B}_{\phi_{t-1}}[\mathbf{T}_{t-1}[\tilde{h}_{0:t-1}] + \tilde{h}_{t-1}(\cdot, x_t)](x_t). \end{aligned}$$

Then, plugging in the particle approximations, we obtain the following recursion

$$\mathbf{T}_t^N[h_{0:t}](x_t) = \sum_{i=1}^N \frac{\omega_{t-1}^i m_t(\xi_{t-1}^i, x_t)}{\sum_{j=1}^N \omega_{t-1}^j m_t(\xi_{t-1}^j, x_t)} \left\{ \mathbf{T}_{t-1}^N[\tilde{h}_{0:t-1}](\xi_{t-1}^i) + \tilde{h}_{t-1}(\xi_{t-1}^i, x_t) \right\}, \quad (5.6)$$

and then

$$\phi_{0:t|t}^N(h_{0:t}) = \sum_{i=1}^N \mathcal{W}_t^i \mathbf{T}_t^N[h_{0:t}](\xi_t^i). \quad (5.7)$$

Equations (5.6)-(5.7) show that  $\mathbf{T}_t^N[h_{0:t}]$  needs only to be estimated at the particle locations and that smoothing estimates for additive functionals can be computed with the forward pass and has  $\mathcal{O}(N^2)$  complexity per time step.

**5.2. Asymptotic variance estimator.** From now on we will assume that  $h_{0:t}$  satisfies (5.5). Our estimator is based on the following alternative expression of the asymptotic variance (5.4)

$$\mathcal{V}_{0:t|t}^{\text{FFBS}}(h) = \sum_{s=0}^t \frac{\gamma_s(\mathbf{1}) \gamma_s(\mathbf{G}_{s,t}[g_t\{h_{0:t} - \phi_{0:t|t}(h_{0:t})\}]^2)}{\gamma_{t+1}(\mathbf{1})^2}, \quad (5.8)$$

which is deduced using the definitions given in Section 3. This expression is motivated by Proposition 5.1 in which we express the numerators that appear in (5.8) in terms of expectations with respect to  $\mathcal{Q}_{e_s, t}$ . The proof is given in Section A.3 of the appendix.

**Proposition 5.1.** *For any  $s \in [0 : t]$  and any additive functional  $h_{0:t} \in \mathbb{F}(\mathcal{X}^{\otimes t+1})$ ,*

$$\gamma_s(\mathbf{1}) \gamma_s(\mathbf{G}_{s,t}[h_{0:t}]^2) = \mathcal{Q}_{e_s, t}([\mathbf{T}_s[\tilde{h}_{0:s}] + \tilde{h}_{s:t}]^{\otimes 2}). \quad (5.9)$$

For every  $t \in \mathbb{N}^*$ , let  $\mathbb{A}_b(\mathcal{X}^{\otimes t+1})$  denote the set of additive functionals (5.5) such that  $\tilde{h}_s \in \mathbb{F}_b(\mathcal{X}^{\otimes 2})$  for all  $s \in [0 : t-1]$ . By Theorem 4.4, for any  $h_{0:t} \in \mathbb{A}_b(\mathcal{X}^{\otimes t+1})$  we have that  $\mathcal{Q}_{e_s,t}^{N,BS}([\mathbf{T}_s[\tilde{h}_{0:s}] + \tilde{h}_{s:t}]^{\otimes 2})$  is a consistent estimator of  $\mathcal{Q}_{e_s,t}([\mathbf{T}_s[h_{0:s}] + \tilde{h}_{s:t}]^{\otimes 2})$ , but  $\mathbf{T}_s[h_{0:s}]$  is intractable and we only have access to its particle approximation  $\mathbf{T}_s^N[h_{0:s}]$ . Our proposed estimator of the asymptotic variance (5.3) is then

$$\mathcal{V}_{0:t|t}^{N,BS}(h_t) := \sum_{s=0}^T \frac{\mathcal{Q}_{e_s,t}^{N,BS}([g_t\{\mathbf{T}_s^N[\tilde{h}_{0:s}] + \tilde{h}_{s:t} - \phi_{0:t|t}^N(h_{0:t})\}]^{\otimes 2})}{\gamma_{t+1}^N(\mathbf{1})^2}, \quad (5.10)$$

where we have replaced  $\phi_{0:t|t}(h_t)$  by its FFBS estimator. Remark that Theorem 4.4 cannot be applied to  $\mathcal{Q}_{e_s,t}^{N,BS}([g_t\{\mathbf{T}_s^N[\tilde{h}_{0:s}] + \tilde{h}_{s:t}]^{\otimes 2})$  because its proof relies on the fact that the function  $h$  integrated by  $\mathcal{Q}_{b,t}^{N,BS}$  does not depend on the particles.

Theorem 5.2 proved in Section A.8 of the appendix shows that weak consistency still holds under the assumptions of Theorem 4.4. The proof proceeds in three steps. We first establish that for all  $s > 0$  and additive functional  $h_{0:s} \in \mathbb{A}_b(\mathcal{X}^{\otimes s+1})$ ,  $\mathbf{T}_s^N[h_{0:s}](x_s)$  converges  $\mathbb{P}$ -a.s. to  $\mathbf{T}_s[h_{0:s}](x_s)$  for any  $x_s \in \mathbf{X}$ . Then, we use it to show that at  $t = s$ , the distance in  $\mathbf{L}_2$  between  $\mathcal{Q}_{e_s,s}^{N,BS}([\mathbf{T}_s^N[h_{0:s}]c_s + \tilde{h}_s] \otimes [\mathbf{T}_s^N[f_{0:s}]d_s + \tilde{f}_s])$  and the "idealized" consistent estimator  $\mathcal{Q}_{e_s,s}^{N,BS}([\mathbf{T}_s[h_{0:s}]c_s + \tilde{h}_s] \otimes [\mathbf{T}_s[f_{0:s}]d_s + \tilde{f}_s])$ , goes to 0. Finally, we extend the result to  $t > s$  by induction, similarly to Theorem 4.4.

**Theorem 5.2.** Assume that (A2 : 4) hold. For any  $t \in \mathbb{N}$ ,  $s \in [0 : t]$ ,  $(\tilde{h}_{s:t}, \tilde{f}_{s:t}) \in \mathbb{F}_b(\mathcal{X}^{\otimes t-s+1})^2$ ,  $(c_t, d_t) \in \mathbb{F}_b(\mathcal{X})^2$  and additive functionnals  $(h_{0:s}, f_{0:s}) \in \mathbb{A}_b(\mathcal{X}^{\otimes s+1})^2$ ,

$$\lim_{N \rightarrow \infty} \|\mathcal{Q}_{e_s,t}^{N,BS}([\mathbf{T}_s^N[h_{0:s}]c_t + \tilde{h}_{s:t}] \otimes [\mathbf{T}_s^N[f_{0:s}]d_t + \tilde{f}_{s:t}]) - \mathcal{Q}_{e_s,t}([\mathbf{T}_s[h_{0:s}]c_t + \tilde{h}_{s:t}] \otimes [\mathbf{T}_s[f_{0:s}]d_t + \tilde{f}_{s:t}])\|_2 = 0, \quad (5.11)$$

and for any additive functional  $h_{0:t} \in \mathbb{A}_b(\mathcal{X}^{\otimes t+1})$ ,  $\mathcal{V}_{0:t|t}^{BS}(h_{0:t})$  converges in probability to  $\mathcal{V}_{0:t|t}^{\text{FFBS}}(h_{0:t})$ .

**5.3. Algorithm for marginal smoothing.** We now provide an algorithm for the case  $h_\ell : x_{0:t} \mapsto h_\ell(x_\ell)$  known as the marginal smoothing problem. For such functions (5.10) is defined for  $t \geq \ell$  and simplifies to

$$\mathcal{V}_{\ell|t}^{N,BS}(h_\ell) := \frac{1}{\gamma_{t+1}^N(\mathbf{1})^2} \left\{ \sum_{s=0}^{\ell} \mathcal{Q}_{e_s,t}^{N,BS}([g_t\{h_\ell - \phi_{\ell|t}^N(h_\ell)\}]^{\otimes 2}) + \sum_{s=\ell+1}^t \mathcal{Q}_{e_s,t}^{N,BS}([g_t\{\mathbf{T}_s^N[h_\ell] - \phi_{\ell|t}^N(h_\ell)\}]^{\otimes 2}) \right\}. \quad (5.12)$$

When  $\ell = t$  we recover the term by term estimator of the filter which is consistent with the fact that  $\phi_{t|t}^N(h) = \phi_t^N(h)$ . Using the bilinearity of  $\mathcal{Q}_{b,t}^{N,BS}$  yields

$$\mathcal{V}_{\ell|t}^{BS}(h_\ell) = R_{1,t}^\ell - \phi_{\ell|t}^N(h_\ell)R_{2,t}^\ell + \phi_{\ell|t}^N(h_\ell)^2 R_t, \quad (5.13)$$

where

$$\begin{cases} R_{\ell,t}^1 &:= \sum_{s=0}^{\ell} \mathcal{Q}_{e_s,t}^{N,BS}([g_t h_\ell]^{\otimes 2}) + \sum_{s=\ell+1}^t \mathcal{Q}_{e_s,t}^{N,BS}([g_t \mathbf{T}_s^N[h_\ell]]^{\otimes 2}), \\ R_{\ell,t}^2 &:= \left\{ \sum_{s=0}^{\ell} \mathcal{Q}_{e_s,t}^{N,BS}(g_t h_\ell \otimes g_t) + \mathcal{Q}_{e_s,t}^{N,BS}(g_t \otimes g_t h_\ell) \right\} \\ &\quad + \left\{ \sum_{s=\ell+1}^t \mathcal{Q}_{e_s,t}^{N,BS}(g_t \mathbf{T}_s^N[h_\ell] \otimes g_t) + \mathcal{Q}_{e_s,t}^{N,BS}(g_t \otimes g_t \mathbf{T}_s^N[h_\ell]) \right\}, \\ R_t &:= \sum_{s=0}^t \mathcal{Q}_{e_s,t}^{N,BS}(g_t^{\otimes 2}). \end{cases}$$

Mirroring (4.25), define for any  $t \in \mathbb{N}$ ,  $n \in [0 : t]$  and  $f_n : x_{0:t}, x'_{0:t} \mapsto f(x_n, x'_n)$  the random variable

$$\mathcal{T}_t^{e_s}[f_n](K_t^1, K_t^2) := \mathbb{E}[\mathbb{I}_{e_s,t}(K_{0:t}^1, K_{0:t}^2) f_n(\xi_n^{K_n^1}, \xi_n^{K_n^2}) | \mathcal{F}_{t-1}^N, K_t^1, K_t^2] \quad (5.14)$$

and also for any  $t \geq \ell$  the random variables

$$S_{\ell,t}^1(K_t^1, K_t^2) = \begin{cases} \sum_{s=0}^{\ell} \mathcal{T}_t^{e_s}[h_{\ell}^{\otimes 2}](K_t^1, K_t^2) + \sum_{s=\ell+1}^t \mathcal{T}_t^{e_s}[\mathbf{T}_s^N[h_{\ell}]^{\otimes 2}](K_t^1, K_t^2) & \text{if } t > \ell, \\ S_{\ell}(K_{\ell}^1, K_{\ell}^2) h_{\ell}^{\otimes 2}(\xi_{\ell}^{K_{\ell}^1}, \xi_{\ell}^{K_{\ell}^2}) & \text{if } t = \ell, \end{cases}$$

$$S_{\ell,t}^2(K_t^1, K_t^2) = \begin{cases} \sum_{s=0}^{\ell} \mathcal{T}_t^{e_s}[h_{\ell}^{\oplus 2}](K_t^1, K_t^2) + \sum_{s=\ell+1}^t \mathcal{T}_t^{e_s}[\mathbf{T}_s^N[h_{\ell}]^{\oplus 2}](K_t^1, K_t^2) & \text{if } t > \ell, \\ S_{\ell}(K_{\ell}^1, K_{\ell}^2) h_{\ell}^{\oplus 2}(\xi_{\ell}^{K_{\ell}^1}, \xi_{\ell}^{K_{\ell}^2}) & \text{if } t = \ell, \end{cases}$$

where for any  $f_{\ell}, f_{\ell}^{\oplus 2} : x_{\ell}, x'_{\ell} \mapsto f_{\ell}(x_{\ell}) + f_{\ell}(x'_{\ell})$  and  $S_{\ell}$  is defined in (4.21). Applying the tower property we get that

$$\mathcal{V}_{\ell|t}^{\text{BS}}(h_t) = N \left( \frac{N}{N-1} \right)^t \sum_{i,j \in [N]^2} \mathcal{W}_t^i \mathcal{W}_t^j \{ S_{\ell,t}^1(i, j) - \phi_{\ell|t}^N(h_{\ell}) S_{\ell,t}^2(i, j) + \phi_{\ell|t}^N(h_{\ell})^2 S_t(i, j) \}.$$

The quantities  $S_{\ell,t+1}^1, S_{\ell,t+1}^2$  may be updated online using the following recursions which are again obtained by applying the tower property

$$S_{\ell,t+1}^1(i, j) := \mathcal{T}_{t+1}^{e_{t+1}}(i, j) \mathbf{T}_{t+1}^N[h_{\ell}](\xi_{t+1}^i) \mathbf{T}_{t+1}^N[h_{\ell}](\xi_{t+1}^j) + \mathbb{1}_{i \neq j} \sum_{m,n \in [N]^2} \beta_{t+1}^{\text{BS}}(i, m) \beta_{t+1}^{\text{BS}}(j, n) S_{\ell,t}^1(m, n), \quad (5.15)$$

and

$$S_{\ell,t+1}^2(i, j) := \mathcal{T}_{t+1}^{e_{t+1}}(i, j) \{ \mathbf{T}_{t+1}^N[h_{\ell}](\xi_{t+1}^i) + \mathbf{T}_{t+1}^N[h_{\ell}](\xi_{t+1}^j) \} + \mathbb{1}_{i \neq j} \sum_{m,n \in [N]^2} \beta_{t+1}^{\text{BS}}(i, m) \beta_{t+1}^{\text{BS}}(j, n) S_{\ell,t}^2(m, n). \quad (5.16)$$

The updates of  $S_{\ell,t+1}^1(i, j)$  and of  $S_{\ell,t+1}^2(i, j)$  are thus similar to that of  $S_t$  in (4.26). The computation of the variance estimator is described in Algorithm 2.

## 6. NUMERICAL SIMULATIONS

We now demonstrate our estimators on particle filtering and smoothing examples in HMMs (see Ex. 3.1). We assume in this section that  $\mathbf{X} = \mathbf{Y} = \mathbb{R}$  and that the dominating measure is the Lebesgue measure. The model considered is the *stochastic volatility model* with, for all  $n \geq 1$ ,

$$\begin{aligned} X_{n+1} &= \varphi X_n + U_{n+1}, \\ Y_n &= \beta \exp(X_n/2) V_n, \end{aligned} \quad (6.1)$$

with  $(\varphi, \beta, \sigma) = (.975, .641, .165)$ ,  $\{U_n\}_{n \in \mathbb{N}}$  and  $\{V_n\}_{n \in \mathbb{N}}$  are two sequences of independent standard Gaussian noises and  $U_n$  is independent of  $V_m$  for all  $(n, m) \in \mathbb{N}^2$ . The state process  $\{X_n\}_{n \in \mathbb{N}}$  is initialized with a Gaussian distribution with zero mean and variance  $\sigma^2/(1 - \phi^2)$ . These are the exact values and initialization used in [32]. The assumptions on the model under which [32] conduct their theoretical analysis and (A1 : 4) are satisfied for this model. All the simulations are run on GPU and the implementations using matrix operations are available at <https://github.com/yazidjanati/asymptoticvariance>.

**Algorithm 2** Variance estimator for marginal smoothing

---

**Require:**  $\mathcal{W}_{t+1}^{1:N}, \mathcal{W}_t^{1:N}, \beta_{t+1}^{\text{BS}}, \mathbf{T}_{t+1}^N[h_\ell], \mathcal{T}_t^0, \mathbf{S}_{\ell,t}^1, \mathbf{S}_{\ell,t}^2, \phi_{\ell|t+1}^N(h_\ell)$

Compute  $\bar{\mathcal{T}}_{t+1}^{e_{t+1}} = \beta_{t+1}^{\text{BS}} \mathcal{T}_t^0 \mathcal{W}_t, \quad \bar{\mathcal{T}}_{t+1}^0 = \beta_{t+1}^{\text{BS}} \mathcal{T}_t^0 \beta_{t+1}^{\text{BS}\top}, \quad \tilde{\mathbf{S}}_{t+1} = \beta_{t+1}^{\text{BS}} \mathbf{S}_t \beta_{t+1}^{\text{BS}\top}$

Set  $\mathcal{T}_{t+1}^{e_{t+1}} = \text{Diag}(\bar{\mathcal{T}}_{t+1}^{e_{t+1}}), \quad \mathcal{T}_{t+1}^0 = \bar{\mathcal{T}}_{t+1}^0 - \text{Diag}(\bar{\mathcal{T}}_{t+1}^0), \quad \mathbf{S}_{t+1} = \tilde{\mathbf{S}}_{t+1} - \text{Diag}(\tilde{\mathbf{S}}_{t+1}) + \mathcal{T}_{t+1}^{e_{t+1}}$

**for**  $i \in \{1, 2\}$  **do**

    Compute  $\tilde{\mathbf{S}}_{\ell,t+1}^i = \beta_{t+1}^{\text{BS}} \mathbf{S}_{\ell,t}^i \beta_{t+1}^{\text{BS}\top}$

    Set  $\tilde{\mathbf{S}}_{\ell,t+1}^i = \tilde{\mathbf{S}}_{\ell,t+1}^i - \text{Diag}(\tilde{\mathbf{S}}_{\ell,t+1}^i)$

**end for**

Set  $\tilde{\mathbf{S}}_{\ell,t+1}^1 = \tilde{\mathbf{S}}_{\ell,t+1}^1 + \mathcal{T}_{t+1}^{e_{t+1}} \odot [\mathbf{T}_{t+1}[h_\ell] \mathbf{T}_{t+1}[h_\ell]^\top]$

Set  $\tilde{\mathbf{S}}_{\ell,t+1}^2 = \tilde{\mathbf{S}}_{\ell,t+1}^2 + \mathcal{T}_{t+1}^{e_{t+1}} \odot [\mathbf{T}_{t+1}[h_\ell] + \mathbf{T}_{t+1}[h_\ell]^\top]$

Set  $\bar{\mathbf{S}}_{\ell,t+1} = \mathbf{S}_{\ell,t+1}^1 - \phi_{\ell|t+1}^N(h_\ell) \mathbf{S}_{\ell,t+1}^2 + \phi_{\ell|t+1}^N(h_\ell)^2 \mathbf{S}_{t+1}$

**return**  $-N^{t+2}/(N-1)^{t+1} \sum_{i,j \in [N]^2} \mathcal{W}_{t+1}^i \mathcal{W}_{t+1}^j \bar{\mathbf{S}}_{\ell,t+1}(i, j), \mathcal{T}_{t+1}^0, \mathbf{S}_{\ell,t+1}^1, \mathbf{S}_{\ell,t+1}^2, \mathbf{S}_{t+1}$ .

---

**6.1. Asymptotic variance of the predictor.** We are interested in the estimation of the asymptotic variance of the predictor  $\eta_t^N(\text{Id})$  at each time step  $t$ . The estimator is given in Section B.1 of the appendix. We use synthetic datasets sampled from (6.1). The real asymptotic variances are intractable and we estimate them by repeating independently a thousand times the computation of each predictor mean  $\eta_t^N(\text{Id})$  with  $N = 10000$  and then multiplying the sample variance by  $N$ .

We next investigate how the three variance estimators based on backward sampling  $\mathcal{V}_{\eta,t}^{N,\text{BS}}, \mathcal{V}_{\eta,t}^{N,M}$  and  $\bar{\mathcal{V}}_{\eta,t}^{N,\text{BS}}$  behave in terms of computational time, bias and variance. The *PaRIS* estimator is used with  $M = 3$  without rejection sampling. For this first experiment, we sampled 750 observations from (6.1) and ran 50 particle filters with  $N = 3000$  from which we obtained 50 replicates of each asymptotic variance estimate. The results are reported in Figure 1. The three estimators exhibit approximately the same variance but the term by term version becomes slightly more biased at  $t$  increases. Strikingly, the *PaRIS* estimator  $\mathcal{V}_{\eta,t}^{N,M}$  behaves similarly to  $\mathcal{V}_{\eta,t}^{N,\text{BS}}$  with a much lower computational time and complexity as can be seen on the left plot of Figure 2.

We now compare  $\mathcal{V}_{\eta,t}^{N,M}$  with  $M = 3$  to [4, 27, 32] on two different observation records of different length. For the lag size parameter  $\lambda$ , we found that  $\lambda = 20$  has the best bias-variance trade-off by comparing the obtained fixed lag estimates with the crude asymptotic variance estimator. Note that in realistic situations choosing the right  $\lambda$  is non trivial (besides the strong mixing case, as argued in [32]) and for this reason we conducted the experiments with two additional lag values,  $\lambda \in \{100, 200\}$ . For moderately long observation records ( $t \in [750]$ ) and with  $N = 3000$ ,  $\mathcal{V}_{\eta,t}^{N,M}$  compares favorably in terms of bias-variance trade-off with the best lagged estimator and even has similar computational time on GPU, as can be seen in Figure 2. The results are reported in Figure 3. The five different estimators are computed with the same particle cloud and replicated 50 times. As expected, when the lag is increased the fixed lag estimates exhibit more variance because of the particle degeneracy. In the extreme case where the lag is set to 750, which corresponds to the CLE, bias and variance both increase significantly, as showcased in the fifth plot of the same figure.

For the longer time horizon  $t \in [3000]$  we set  $N = 5000$  and picked three time steps in order to monitor the bias and variance closely, see Figure 4. The variance of  $\mathcal{V}_{\eta,t}^{N,M}$  remains steady while the bias increases gradually but slowly. Nonetheless, our estimator remains competitive with the

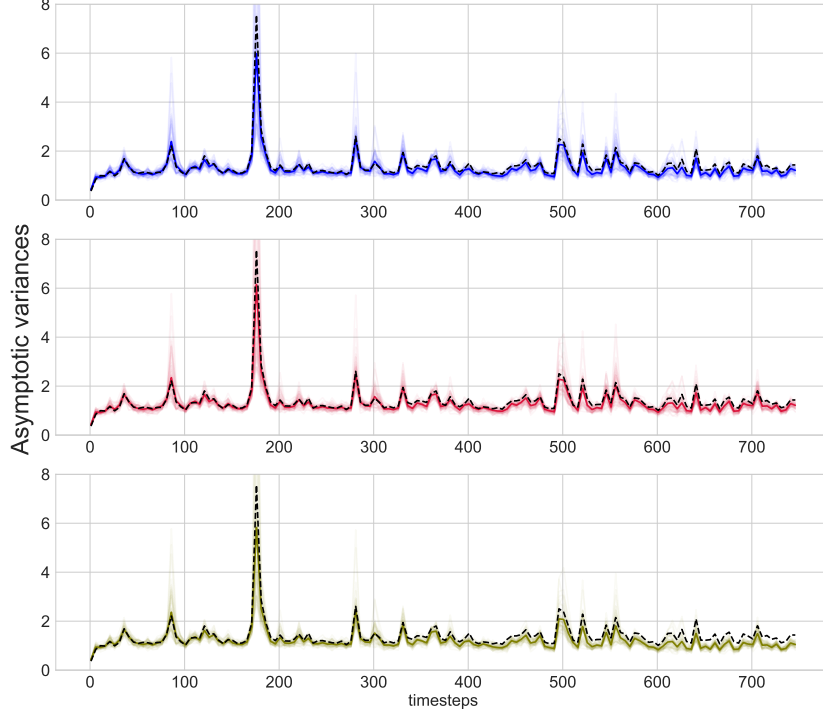


FIGURE 1. Long-term behavior of  $\mathcal{V}_{\eta,t}^{N,BS}$  (top),  $V_{\eta,t}^{N,M}$  with  $M = 3$  (middle) and  $\bar{V}_{\eta,t}^{N,BS}$  (bottom). The black dashed line is the asymptotic variance estimated using brute force. The number of particles is set to  $N = 3000$ .

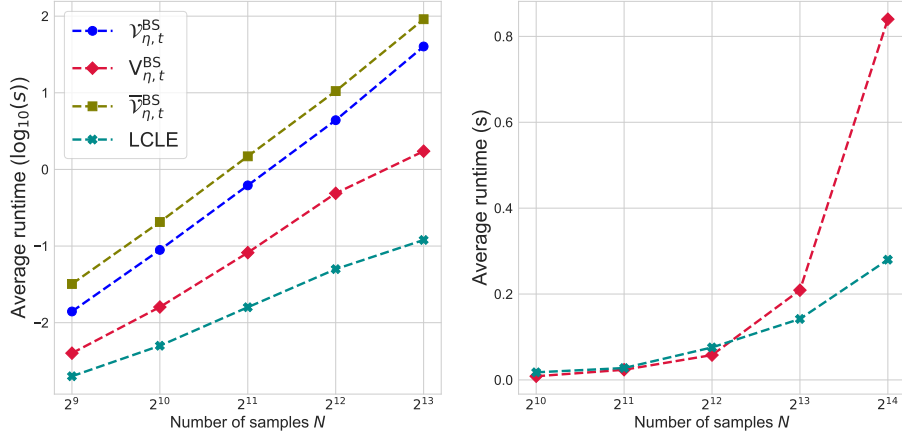


FIGURE 2. Comparison of time complexity (left) and runtime (right) for the different estimators, per time step. The runtime on the left plot is on CPU and that on the right plot on GPU, only for our most competitive estimator.

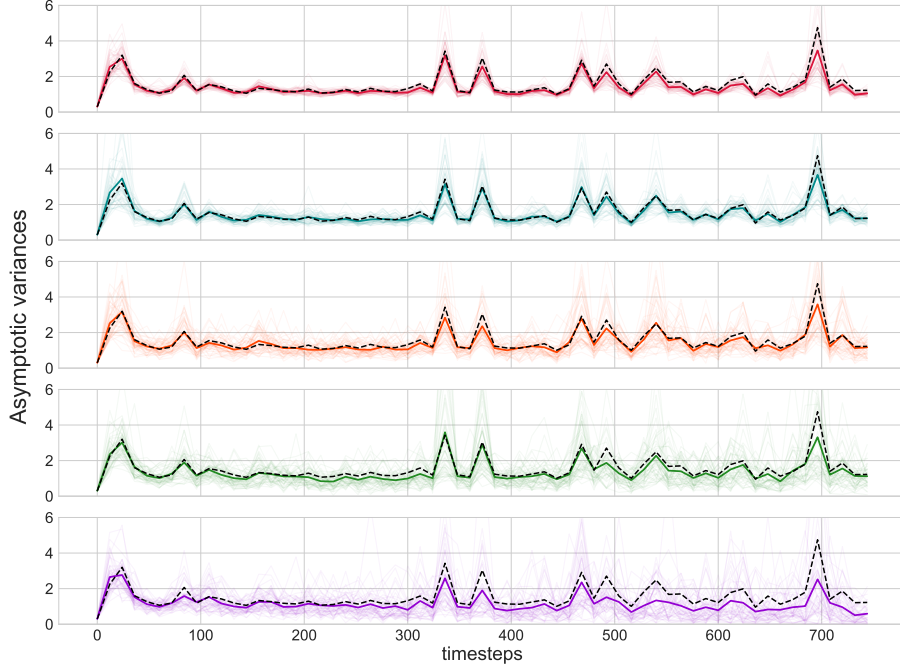


FIGURE 3. Long-term behavior of the asymptotic variance estimators up to  $t = 750$ . From top to bottom: *PaRIS* version of  $\mathcal{V}_{\eta,t}^{N,BS}$  with  $M = 3$ , lagged estimators with (in order)  $\lambda \in \{20, 100, 200, 750\}$ . The case  $\lambda = 750$  corresponds to the CLE estimator. For each estimator, the blurred colored lines represent each run out of fifty runs and solid colored lines correspond to their average. The black dashed line is the asymptotic variance obtained by brute force. The number of particles  $N$  is set to 3000.

best fixed lag estimator. Doubling the number of particles decreases both bias and variance, as highlighted by the right plot.

The computational cost of  $\mathcal{V}_{\eta,t}^{N,M}$  is approximately twice larger than that of the genealogy tracing estimators when  $N = 5000$  as shown in Figure 2. However, we are able to maintain a small variance and bias without having to tune any other parameter besides the sample size. In more complicated models or realistic scenarios, one does not know in advance which lag size is suitable and using an inappropriate lag might yield poor estimates.

Let us now explain the slight bias observed in our estimators. Assume for the sake of illustration that the transition kernels are strongly mixing (Assumptions (A4 : 5)) and that there exist  $0 < G_- < G_\infty < \infty$  such that for all  $x \in \mathcal{X}$ ,  $G_- \leq g_t(x) \leq G_\infty$ . Set  $\rho := \sigma_+ G_\infty / \sigma_- G_-$ . Then, for any  $(i, j) \in [N]^2$ ,  $\beta_t^{BS}(i, j) \leq \rho/N$  and for  $N > \rho^2$ ,

$$\mathcal{T}_t^0(k_t^1, k_t^2) \geq (1 - \rho^2/N)^t, \quad (6.2)$$

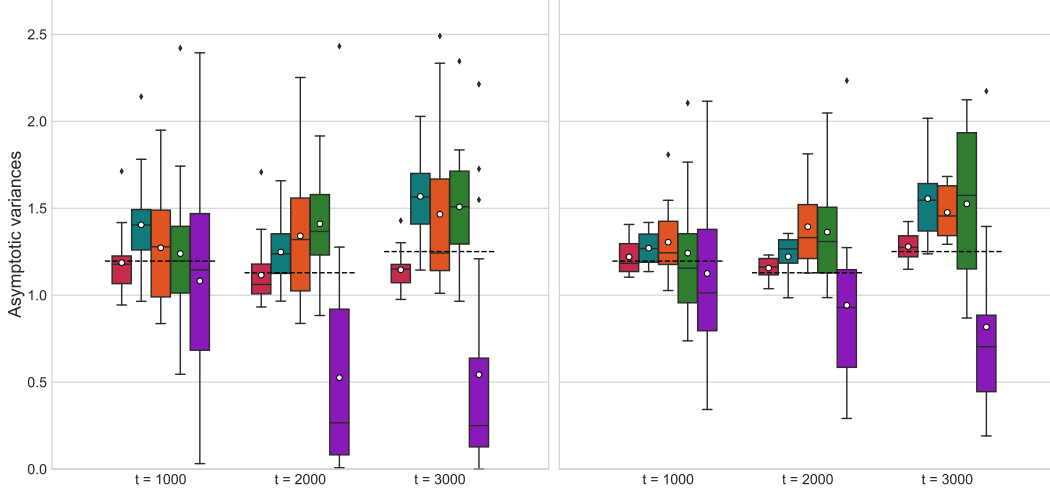


FIGURE 4. Boxplots of the long-term behavior of the asymptotic variance estimates up to  $t = 3000$ . VBS refers to  $V_{\eta,t}^{N,M}$  with  $M = 3$ . The white dots represent the average of the asymptotic variance estimates of each algorithm at the specified time  $t$ . The dashed black lines correspond to the asymptotic variances estimated by brute force.  $N$  is set to 5000 on the left boxplot and 10000 on the right one. The boxplots at each time step from left to right are:  $\tilde{Q}_{\eta,t}^{N,M}$  with  $M = 3$  and then the lagged CLEs with  $\lambda \in \{20, 100, 200, 3000\}$ , in order.

by using repeatedly the fact that

$$\sum_{k_{t-1}^1: k_{t-1}^2 \in [N]^2} \mathbb{1}_{k_{t-1}^1 \neq k_{t-1}^2} \beta_t^{\text{BS}}(k_t^1, k_{t-1}^1) \beta_t^{\text{BS}}(k_t^2, k_{t-1}^2) = 1 - \sum_{i=1}^N \beta_t^{\text{BS}}(k_t^1, i) \beta_t^{\text{BS}}(k_t^2, i) \geq 1 - \rho^2/N.$$

The lowerbound (6.2) sheds light on what should be expected from our estimator. For the variance estimator to maintain low bias,  $\mathcal{T}_t^0$  should stay relatively large, i.e. be above some threshold  $\epsilon > 0.5$  (this is the value observed in our experiments), which means that  $N$  should be larger than  $\rho^2/(1 - \epsilon^{1/t})$  by (6.2). As to (3.11), although it is expected to collapse to 0 after  $\mathcal{O}(N)$  timesteps following the findings of [25], the estimator starts to exhibit high bias and variance much before the collapse as the set of time 0 ancestors depletes.

**6.2. Asymptotic variance of the smoother.** Here we are interested in the estimation of the asymptotic variance associated to the FFBS estimates of the marginal means  $\phi_{\ell|t}(\text{Id})$  with  $\ell$  fixed and  $t \geq \ell$  varying using Algorithm 2. For this example we sampled four different observation records of length 160 each and  $\ell$  is set to 100. The real asymptotic variances of each  $\phi_{\ell|t}(\text{Id})$  are intractable and they are estimated using 1000 independent replicates of the marginal means  $\phi_{\ell|t}^N(\text{Id})$  with  $N = 10000$ . We then multiply the obtained sample variance by  $N$ . The results are reported in Figure 5. As expected, the crude estimates of the asymptotic variances all stagnate after some time  $t$  due to the incoming observations becoming less and less informative as  $t$  grows and thus no longer influencing the value of  $\phi_{\ell|t}(\text{Id})$ .

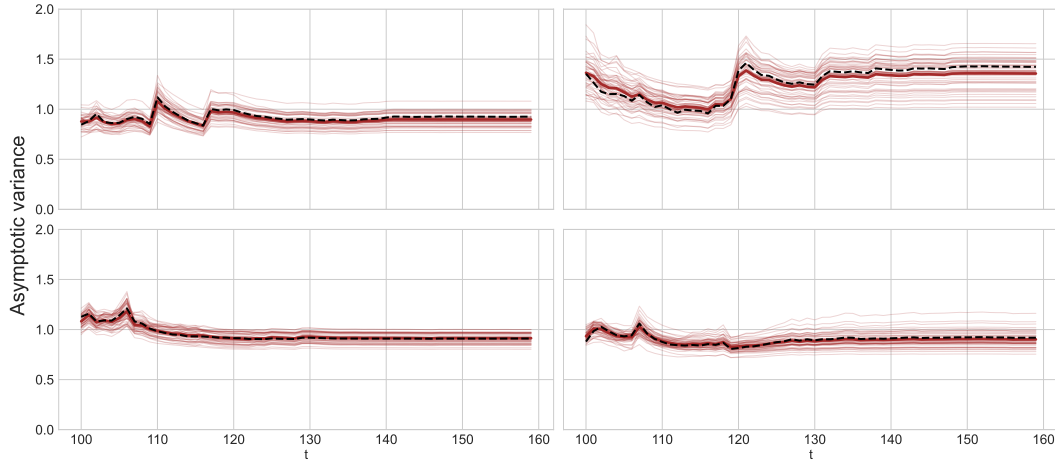


FIGURE 5. Asymptotic variance estimates for four different observation records of the marginal mean  $\phi_{100|t}^N(\text{Id})$  where  $t \in [100, 160]$ . The blurred brown lines on the left plot represent 50 runs and the solid brown line their average. The black dashed line is the crude variance estimator. The number of particles  $N$  is set to 5000.

The estimator proposed in 5.3 captures well the behavior of the asymptotic variance with little variance and also stagnates at the same time. We observed in our experiments that, in comparison with the variance estimators of filtering algorithms, for this estimator to provide good performance more samples are required for shorter time horizons. This could be explained by the additional level of approximation coming from  $\mathbf{T}_s^N$ . Nonetheless, the increased computational time incurred by the increase of the number of particles is to be compared with the time it takes to compute the crude variance estimates up to  $t = 160$ , which is about 1 hour when running on GPU. In comparison, our estimator takes 3 minutes.

#### REFERENCES

- [1] Christophe Andrieu, Arnaud Doucet, and Roman Holenstein. Particle Markov chain Monte Carlo methods. *Journal of the Royal Statistical Society: Series B*, 72(3):269–342, 2010.
- [2] Christophe Andrieu, Anthony Lee, and Matti Vihola. Uniform ergodicity of the iterated conditional SMC and geometric ergodicity of particle Gibbs samplers. *Bernoulli*, 24(2):842 – 872, 2018.
- [3] Olivier Cappé, Eric Moulines, and Tobias Ryden. *Inference in Hidden Markov Models (Springer Series in Statistics)*. Springer-Verlag, Berlin, Heidelberg, 2005.
- [4] Hock Peng Chan and Tze Leung Lai. A general theory of particle filters in hidden Markov models and some applications. *The Annals of Statistics*, 41(6):2877 – 2904, 2013.
- [5] Nicolas Chopin. Central limit theorem for sequential monte carlo methods and its application to bayesian inference. *The Annals of Statistics*, 32(6):2385–2411, 2004.
- [6] Nicolas Chopin and Omiros Papaspiliopoulos. *An Introduction to Sequential Monte Carlo*. Springer International Publishing, 2020.



- [7] Nicolas Chopin and Sumeetpal S. Singh. On particle Gibbs sampling. *Bernoulli*, 21(3):1855 – 1883, 2015.
- [8] F. Cérou, P. Del Moral, and A. Guyader. A nonasymptotic theorem for unnormalized Feynman–Kac particle models. *Annales de l’Institut Henri Poincaré, Probabilités et Statistiques*, 47(3):629 – 649, 2011.
- [9] Pierre Del Moral. Feynman-kac formulae. In *Feynman-Kac Formulae*, pages 47–93. Springer, 2004.
- [10] Pierre Del Moral, Arnaud Doucet, and Sumeetpal Singh. Forward smoothing using sequential monte carlo. *arXiv preprint arXiv:1012.5390*, 2010.
- [11] Pierre Del Moral, Arnaud Doucet, and Sumeetpal S Singh. A backward particle interpretation of feynman-kac formulae. *ESAIM: Mathematical Modelling and Numerical Analysis*, 44(5):947–975, 2010.
- [12] Pierre Del Moral, Arnaud Doucet, and Singh Sumeetpal. Forward smoothing using sequential monte carlo. *ArXiv:1012.5390*, 2010.
- [13] Pierre Del Moral and Alice Guionnet. Central limit theorem for nonlinear filtering and interacting particle systems. *Annals of Applied Probability*, pages 275–297, 1999.
- [14] Randal Douc and Olivier Cappé. Comparison of resampling schemes for particle filtering. In *ISPA 2005. Proceedings of the 4th International Symposium on Image and Signal Processing and Analysis, 2005.*, pages 64–69. IEEE, 2005.
- [15] Randal Douc, Aurélien Garivier, Eric Moulines, and Jimmy Olsson. Sequential Monte Carlo smoothing for general state space hidden Markov models. *The Annals of Applied Probability*, 21(6):2109 – 2145, 2011.
- [16] Randal Douc, Eric Moulines, and David Stoffer. *Nonlinear time series: Theory, methods and applications with R examples*. CRC press, 2014.
- [17] Arnaud Doucet, Simon Godsill, and Christophe Andrieu. On sequential monte carlo sampling methods for bayesian filtering. *Statistics and computing*, 10(3):197–208, 2000.
- [18] Qiming Du and Arnaud Guyader. Variance estimation in adaptive sequential Monte Carlo. *The Annals of Applied Probability*, 31(3):1021 – 1060, 2021.
- [19] Cyrille Durrty and Sylvain Le Corff. Non-asymptotic deviation inequalities for smoothed additive functionals in nonlinear state-space models. *Bernoulli*, 19(5B):2222 – 2249, 2013.
- [20] Paul Fearnhead, David Wyncoll, and Jonathan Tawn. A sequential smoothing algorithm with linear computational cost. *Biometrika*, 97(2):447–464, 2010.
- [21] Pierre Gloaguen, Sylvain Le Corff, and Jimmy Olsson. A pseudo-marginal sequential monte carlo online smoothing algorithm. *Bernoulli*, to appear, 2022.
- [22] Simon J Godsill, Arnaud Doucet, and Mike West. Monte carlo smoothing for nonlinear time series. *Journal of the american statistical association*, 99(465):156–168, 2004.
- [23] Neil J Gordon, David J Salmond, and Adrian FM Smith. Novel approach to nonlinear/non-gaussian bayesian state estimation. In *IEE Proceedings F-radar and signal processing*, volume 140, pages 107–113. IET, 1993.
- [24] Nikolas Kantas, Arnaud Doucet, Sumeetpal S. Singh, Jan Maciejowski, and Nicolas Chopin. On particle methods for parameter estimation in state-space models. *Statist. Sci.*, 30(3):328–351, August 2015.
- [25] Jere Koskela, Paul A Jenkins, Adam M Johansen, and Dario Spano. Asymptotic genealogies of interacting particle systems with an application to sequential monte carlo. *The Annals of Statistics*, 48(1):560–583, 2020.

- [26] Hans R Künsch. Recursive monte carlo filters: algorithms and theoretical analysis. *The Annals of Statistics*, 33(5):1983–2021, 2005.
- [27] A Lee and N Whiteley. Variance estimation in the particle filter. *Biometrika*, 105(3):609–625, 06 2018.
- [28] Anthony Lee, Sumeetpal S Singh, and Matti Vihola. Coupled conditional backward sampling particle filter. *The Annals of Statistics*, 48(5):3066–3089, 2020.
- [29] Fredrik Lindsten and Thomas B. Schön. On the use of backward simulation in the particle gibbs sampler. In *2012 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 3845–3848, 2012.
- [30] Jane Liu and Mike West. Sequential monte carlo methods in practice. *Statistics for Engineering and Information Science*, edited by A. Doucet, N. Freitas, and N. Gordon (Springer, New York, 2001), pages 225–246, 2001.
- [31] Jimmy Olsson, Olivier Cappé, Randal Douc, and Eric Moulines. Sequential monte carlo smoothing with application to parameter estimation in nonlinear state space models. *Bernoulli*, 14(1):155–179, 2008.
- [32] Jimmy Olsson and Randal Douc. Numerically stable online estimation of variance in particle filters. *Bernoulli*, 25(2):1504 – 1535, 2019.
- [33] Jimmy Olsson and Johan Westerborn. Efficient particle-based online smoothing in general hidden Markov models: The PaRIS algorithm. *Bernoulli*, 23(3):1951 – 1996, 2017.
- [34] Michael K Pitt and Neil Shephard. Filtering via simulation: Auxiliary particle filters. *Journal of the American statistical association*, 94(446):590–599, 1999.
- [35] George Poyiadjis, Arnaud Doucet, and Sumeetpal 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, 2011.
- [36] Halsey Lawrence Royden and Patrick Fitzpatrick. *Real analysis*, volume 32. Macmillan New York, 1988.
- [37] Hisashi Tanizaki and Roberto Mariano. Prediction, filtering and smoothing in non-linear and non-normal cases using monte carlo integration. *Journal of Applied Econometrics*, 9(2):163–79, 1994.

## APPENDIX A. PROOFS

In this section we provide the proofs of Propositions 4.3, 5.1 and Theorems 4.4, 4.7, 4.8 and 5.2. The various Propositions and Lemmata used are stated and proved in Section A.9-A.10. Other intermediary technical results are provided in Section C.

**A.1. Preliminaries.** In order to simplify the notations, in what follows we write

$$\mathcal{Q}_{b,t}^{N,\text{BS}}(h) = \sum_{k_{0:t}^{1:2} \in [N]^{2(t+1)}} \bar{\Lambda}_{b,t}^{1,2}(k_{0:t}^1, k_{0:t}^2) h(\xi_{0:t}^{k_{0:t}^1}, \xi_{0:t}^{k_{0:t}^2}),$$

where

$$\bar{\Lambda}_{b,t}^{1,2}(k_{0:t}^1, k_{0:t}^2) := \prod_{s=0}^t N^{b_s} \left( \frac{N}{N-1} \right)^{1-b_s} \gamma_t^N(\mathbf{1})^2 \mathbb{I}_{b,t}(k_{0:t}^1, k_{0:t}^2) \Lambda_{1,t}^{\text{BS}}(k_{0:t}^1) \Lambda_{2,t}^{\text{BS}}(k_{0:t}^1; k_{0:t}^2). \quad (\text{A1.1})$$

By (4.13),

$$\begin{aligned} \Lambda_{1,t}^{\text{BS}}(k_{0:t}^1) &= \beta_t^{\text{BS}}(k_t^1, k_{t-1}^1) \Lambda_{1,t-1}^{\text{BS}}(k_{0:t-1}^1), \\ \Lambda_{2,t}^{\text{BS}}(k_{0:t}^1; k_{0:t}^2) &= \{ \beta_t^{\text{BS}}(k_t^2, k_{t-1}^2) \mathbb{1}_{k_t^2 \neq k_t^1} + \mathcal{W}_{t-1}^{k_t^2-1} \mathbb{1}_{k_t^2 = k_t^1} \} \Lambda_{2,t-1}^{\text{BS}}(k_{0:t-1}^1; k_{0:t-1}^2). \end{aligned}$$

and by (3.4) we have that  $\gamma_t^N(\mathbf{1}) = \gamma_{t-1}^N(\mathbf{1}) N^{-1} \Omega_{t-1}$ , hence, using (4.16)  $\mathcal{Q}_{b,t}^{N,\text{BS}}(h)$  becomes

$$\begin{aligned} \mathcal{Q}_{b,t}^{N,\text{BS}}(h) &= \sum_{k_{0:t}^{1:2} \in [N]^{2(t+1)}} \bar{\Lambda}_{b,t}^{1,2}(k_{0:t}^1, k_{0:t}^2) h(\xi_{0:t}^{k_{0:t}^1}, \xi_{0:t}^{k_{0:t}^2}) \quad (\text{A1.2}) \\ &= \sum_{k_{0:t-1}^{1:2} \in [N]^{2t}} \bar{\Lambda}_{b,t-1}^{1,2}(k_{0:t-1}^1, k_{0:t-1}^2) \frac{\Omega_{t-1}^2}{N^2} \sum_{k_t^{1:2} \in [N]^2} \beta_t^{\text{BS}}(k_t^1, k_{t-1}^1) \\ &\quad \times \left[ \frac{N}{N-1} \beta_t^{\text{BS}}(k_t^2, k_{t-1}^2) \mathbb{1}_{k_t^2 \neq k_t^1, b_t=0} + N \mathcal{W}_{t-1}^{k_t^2-1} \mathbb{1}_{k_t^2 = k_t^1, b_t=1} \right] h(\xi_{0:t}^{k_{0:t}^1}, \xi_{0:t}^{k_{0:t}^2}). \end{aligned}$$

## A.2. Proof of Proposition 4.3.

*Proof of Lemma 4.1.* By (A1), for any  $(x, y) \in \mathbb{X}^2$ ,

$$\begin{aligned} \beta_t^N(x, y) \phi_{t-1}^N M_t(dx) &= \beta_t^N(x, y) \sum_{i=1}^N \mathcal{W}_{t-1}^i M_t(\xi_{t-1}^i, dx) \\ &= \frac{g_{t-1}(y) m_t(y, x)}{\sum_{i=1}^N \omega_{t-1}^i m_t(\xi_{t-1}^i, x)} \sum_{i=1}^N \mathcal{W}_{t-1}^i m_t(\xi_{t-1}^i, x) \nu(dx) \\ &= \frac{g_{t-1}(y)}{\Omega_{t-1}} M_t(y, dx). \end{aligned}$$

Consequently, for any  $(k_{t-1}^1, k_t^1) \in \mathbb{X}^2$  and  $h \in \mathbb{F}(\mathcal{X})$ ,

$$\begin{aligned} \mathbb{E}[\beta_t^N(\xi_t^{k_t}, \xi_{t-1}^{k_{t-1}}) h(\xi_t^{k_t}) | \mathcal{F}_{t-1}^N] &= \int \beta_t^N(x, \xi_{t-1}^{k_{t-1}}) h(x) \phi_{t-1}^N M_t(dx) \\ &= \int \frac{1}{\Omega_{t-1}} \omega_{t-1}^{k_{t-1}^1} h(x) M_t(\xi_{t-1}^{k_{t-1}^1}, dx) \\ &= \mathcal{W}_{t-1}^{k_{t-1}^1} M_t[h](\xi_{t-1}^{k_{t-1}^1}). \end{aligned}$$

On the other hand,

$$\mathbb{E}[\mathbb{1}_{k_{t-1}=A_{t-1}^{k_t}} h(\xi_t^{k_t}) | \mathcal{F}_{t-1}^N] = \int \sum_{i=1}^N \mathbb{1}_{k_{t-1}=i} \mathcal{W}_{t-1}^i M_t(\xi_{t-1}^i, dx) h(x) = \mathcal{W}_{t-1}^{k_{t-1}} M_t[h](\xi_{t-1}^{k_{t-1}}) .$$

□

*Proof of Proposition 4.3.* For the proof of i), note that conditionally on  $\mathcal{F}_{t-1}^N$ ,  $\xi_t^{1:N}$  are i.i.d. with distribution  $\phi_{t-1}^N M_t$ , hence,

$$\begin{aligned} & \mathbb{E} \left[ \sum_{k_t^{1:2} \in [N]^2} \mathbb{1}_{k_t^1 \neq k_t^2} \beta_t^{\text{BS}}(k_t^1, k_{t-1}^1) \beta_t^{\text{BS}}(k_t^2, k_{t-1}^2) h(\xi_{0:t}^{k_{0:t}^1}, \xi_{0:t}^{k_{0:t}^2}) \middle| \mathcal{F}_{t-1}^N \right] \\ &= \sum_{k_t^{1:2} \in [N]^2} \int \mathbb{1}_{k_t^1 \neq k_t^2} \beta_t^{\text{BS}}(k_t^1, k_{t-1}^1) \beta_t^{\text{BS}}(k_t^2, k_{t-1}^2) h(\xi_{0:t}^{k_{0:t}^1}, \xi_{0:t}^{k_{0:t}^2}) \phi_{t-1}^N M_t(d\xi_t^{k_t^1}) \phi_{t-1}^N M_t(d\xi_t^{k_t^2}) , \end{aligned}$$

and by (4.11) in Lemma 4.1,

$$\begin{aligned} & \mathbb{E} \left[ \sum_{k_t^{1:2} \in [N]^2} \mathbb{1}_{k_t^1 \neq k_t^2} \beta_t^{\text{BS}}(k_t^1, k_{t-1}^1) \beta_t^{\text{BS}}(k_t^2, k_{t-1}^2) h(\xi_{0:t}^{k_{0:t}^1}, \xi_{0:t}^{k_{0:t}^2}) \middle| \mathcal{F}_{t-1}^N \right] \\ &= \frac{N(N-1)}{\Omega_{t-1}^2} (g_{t-1}^{\otimes 2} \mathcal{M}_t^0[h]) (\xi_{0:t-1}^{k_{0:t-1}^1}, \xi_{0:t-1}^{k_{0:t-1}^2}) , \end{aligned} \quad (\text{A1.3})$$

where

$$g_{t-1}^{\otimes 2} \mathcal{M}_t^0[h] : (x_{0:t-1}, x'_{0:t-1}) \mapsto g_{t-1}^{\otimes 2}(x_{t-1}, x'_{t-1}) \int h(x_{0:t}, x'_{0:t}) M_t(x_{t-1}, dx_t) M_t(x'_{t-1}, dx'_t) .$$

On the other hand, by (4.11) in Lemma 4.1,

$$\begin{aligned} & \mathbb{E} \left[ \sum_{k_t^{1:2} \in [N]^2} \mathbb{1}_{k_t^1 = k_t^2} \beta_t^{\text{BS}}(k_t^1, k_{t-1}^1) \mathcal{W}_{t-1}^{k_{t-1}^2} h(\xi_{0:t}^{k_{0:t}^1}, \xi_{0:t}^{k_{0:t}^2}) \middle| \mathcal{F}_{t-1}^N \right] \\ &= \sum_{k_t^{1:2} \in [N]^2} \int \mathbb{1}_{k_t^1 = k_t^2} \beta_t^{\text{BS}}(k_t^1, k_{t-1}^1) \mathcal{W}_{t-1}^{k_{t-1}^2} h(\xi_{0:t}^{k_{0:t}^1}, \xi_{0:t}^{k_{0:t}^2}) \phi_{t-1}^N M_t(d\xi_t^{k_t^1}) \delta_{\xi_t^{k_t^1}}(d\xi_t^{k_t^2}) \\ &= \frac{1}{\Omega_{t-1}^2} \sum_{k_t^{1:2} \in [N]^2} \mathbb{1}_{k_t^1 = k_t^2} \omega_{t-1}^{k_{t-1}^1} \omega_{t-1}^{k_{t-1}^2} \int h(\xi_{0:t}^{k_{0:t}^1}, \xi_{0:t}^{k_{0:t}^2}) M_t(\xi_{t-1}^{k_{t-1}^1}, d\xi_t^{k_t^1}) \delta_{\xi_t^{k_t^1}}(d\xi_t^{k_t^2}) \\ &= \frac{N}{\Omega_{t-1}^2} (g_{t-1}^{\otimes 2} \mathcal{M}_t^1[h]) (\xi_{0:t}^{k_{0:t}^1}, \xi_{0:t}^{k_{0:t}^2}) , \end{aligned} \quad (\text{A1.4})$$

where

$$g_{t-1}^{\otimes 2} \mathcal{M}_t^1[h] : (x_{0:t-1}, x'_{0:t-1}) \mapsto g_{t-1}^{\otimes 2}(x_{t-1}, x'_{t-1}) \int h(x_{0:t}, x'_{0:t}) M_t(x_{t-1}, dx_t) \delta_{x_t}(dx'_t) .$$

If  $b_t = 0$ , since  $\Omega_{t-1}$  and  $\bar{\Lambda}_{b,t-1}^{1,2}(k_{0:t-1}^1, k_{0:t-1}^2)$  are  $\mathcal{F}_{t-1}^N$  measurable for any  $(k_{0:t-1}^1, k_{0:t-1}^2) \in [N]^{2t}$ , by (A1.2) and (A1.3),

$$\begin{aligned} \mathbb{E} \left[ \mathcal{Q}_{b,t}^{N,BS}(h) | \mathcal{F}_{t-1}^N \right] &= \sum_{k_{0:t-1}^{1,2} \in [N]^2} \bar{\Lambda}_{b,t-1}^{1,2}(k_{0:t-1}^1, k_{0:t-1}^2) \frac{\Omega_{t-1}^2}{N(N-1)} \\ &\quad \times \mathbb{E} \left[ \sum_{k_t^{1,2} \in [N]^2} \mathbb{1}_{k_t^1 \neq k_t^2} \beta_t^{BS}(k_t^1, k_{t-1}^1) \beta_t^{BS}(k_t^2, k_{t-1}^2) h(\xi_{0:t}^{k_{0:t-1}^1}, \xi_{0:t}^{k_{0:t-1}^2}) \middle| \mathcal{F}_{t-1}^N \right] \\ &= \sum_{k_{0:t-1}^{1,2} \in [N]^2} \bar{\Lambda}_{b,t-1}^{1,2}(k_{0:t-1}^1, k_{0:t-1}^2) (g_{t-1}^{\otimes 2} \mathcal{M}_t^0[h]) (\xi_{0:t-1}^{k_{0:t-1}^1}, \xi_{0:t-1}^{k_{0:t-1}^2}) \\ &= \mathcal{Q}_{b,t-1}^{N,BS}(g_{t-1}^{\otimes 2} \mathcal{M}_t^0[h]) . \end{aligned}$$

If  $b_t = 1$ , again by (A1.2) and (A1.4),

$$\begin{aligned} \mathbb{E} \left[ \mathcal{Q}_{b,t}^{N,BS}(h) | \mathcal{F}_{t-1}^N \right] &= \sum_{k_{0:t-1}^{1,2} \in [N]^2} \bar{\Lambda}_{b,t-1}^{1,2}(k_{0:t-1}^1, k_{0:t-1}^2) \frac{\Omega_{t-1}^2}{N} \\ &\quad \times \mathbb{E} \left[ \sum_{k_t^{1,2} \in [N]^2} \mathbb{1}_{k_t^1 = k_t^2} \beta_t^{BS}(k_t^1, k_{t-1}^1) \mathcal{W}_{t-1}^{k_{t-1}^2} h(\xi_{0:t}^{k_{0:t-1}^1}, \xi_{0:t}^{k_{0:t-1}^2}) \middle| \mathcal{F}_{t-1}^N \right] \\ &= \sum_{k_{0:t-1}^{1,2} \in [N]^2} \bar{\Lambda}_{b,t-1}^{1,2}(k_{0:t-1}^1, k_{0:t-1}^2) (g_{t-1}^{\otimes 2} \mathcal{M}_t^1[h]) (\xi_{0:t-1}^{k_{0:t-1}^1}, \xi_{0:t-1}^{k_{0:t-1}^2}) \\ &= \mathcal{Q}_{b,t-1}^{N,BS}(g_{t-1}^{\otimes 2} \mathcal{M}_t^1[h]) . \end{aligned}$$

For the proof of ii), we proceed by induction. Let  $t = 0$  and  $h \in \mathbb{F}(\mathcal{X}^{\otimes 2})$ . If  $b_0 = 0$ , since  $\xi_0^{1:N} \stackrel{\text{iid}}{\sim} M_0$ ,

$$\begin{aligned} \mathbb{E} \left[ \mathcal{Q}_{0,0}^{N,BS}(h) \right] &= \mathbb{E} \left[ \frac{1}{N(N-1)} \sum_{i,j \in [N]^2} \mathbb{1}_{i \neq j} h(\xi_0^i, \xi_0^j) \right] \\ &= \frac{1}{N(N-1)} \sum_{i,j \in [N]^2} \mathbb{1}_{i \neq j} \mathcal{M}_0^0[h] = \mathcal{M}_0^0[h] = \mathcal{Q}_{b,0}(h) . \end{aligned}$$

If  $b_0 = 1$ ,

$$\mathbb{E} \left[ \mathcal{Q}_{1,0}^{N,BS}(h) \right] = \mathbb{E} \left[ \frac{1}{N} \sum_{i=1}^N h(\xi_0^i, \xi_0^i) \right] = \mathcal{M}_0^1[h] = \mathcal{Q}_{b,0}(h) .$$

Let  $t \in \mathbb{N}^*$  and  $h \in \mathbb{F}(\mathcal{X}^{\otimes 2(t+1)})$ . Assume that  $\mathbb{E}[\mathcal{Q}_{b,t-1}^{N,BS}(f)] = \mathcal{Q}_{b,t-1}(f)$  for any  $b \in \mathcal{B}_t$  and  $f \in \mathbb{F}(\mathcal{X}^{\otimes 2t})$ . By (i) in Proposition 4.3 and the tower property

$$\mathbb{E} \left[ \mathcal{Q}_{b,t}^{N,BS}(h) \right] = \mathbb{E} \left[ \mathbb{E} \left[ \mathcal{Q}_{b,t}^{N,BS}(h) | \mathcal{F}_{t-1}^N \right] \right] = \mathbb{E} \left[ \mathcal{Q}_{b,t-1}^{N,BS}(g_{t-1}^{\otimes 2} \mathcal{M}_t^{b_t}[h]) \right]$$

and  $g_{t-1}^{\otimes 2} \mathcal{M}_t^{b_t}[h] \in \mathbb{F}(\mathcal{X}^{\otimes 2t})$ . Thus, by the induction hypothesis and the definition of  $\mathcal{Q}_{b,t-1}$  we get

$$\mathbb{E} \left[ \mathcal{Q}_{b,t}^{N,BS}(h) \right] = \mathbb{E} \left[ \mathcal{Q}_{b,t-1}^{N,BS}(g_{t-1}^{\otimes 2} \mathcal{M}_t^{b_t}[h]) \right] = \mathcal{Q}_{b,t-1}(g_{t-1}^{\otimes 2} \mathcal{M}_t^{b_t}[h]) = \mathcal{Q}_{b,t}(h) ,$$

which completes the proof. The proof of iii) is a direct consequence of (4.4), (4.17) and ii).  $\square$

A.3. **Proof of Proposition 5.1.** Let  $h_{0:t}$  be an additive functional (5.5). By definition, for  $s \in [t]$ ,

$$\begin{aligned} \mathbf{G}_{s,t}(x_s, h_{0:t}) &= \int \{ \tilde{h}_{0:s}(x_{0:s}) + \tilde{h}_{s:t}(x_{s:t}) \} \mathbf{T}_s(x_s, dx_{0:s-1}) \mathbf{Q}_{s+1:t}(x_s, dx_{s+1:t}) \\ &= \int \{ \mathbf{T}_s[\tilde{h}_{0:s}](x_s) + \tilde{h}_{s:t}(x_{s:t}) \} \mathbf{Q}_{s+1:t}(x_s, dx_{s+1:t}) , \end{aligned}$$

and then, setting  $\mathbf{H}_{s:t} : x_{s:t} \mapsto \mathbf{T}_s[\tilde{h}_{0:s}](x_s) + \tilde{h}_{s:t}(x_{s:t})$  we get

$$\begin{aligned} &\gamma_s(\mathbf{1})\gamma_s(\mathbf{G}_{s,t}[h_{0:t}]^2) \\ &= \gamma_s(\mathbf{1})\gamma_s(\mathbf{Q}_{s+1:t}[\mathbf{T}_s[\tilde{h}_{0:s}] + \tilde{h}_{s:t}]^2) \\ &= \int \gamma_{0:s-1}(dx'_{0:s-1}) g_{s-1}(x'_{s-1}) \int \mathbf{Q}_{s+1:t}[\mathbf{H}_{s:t}](x_s) \mathbf{Q}_{s+1:t}[\mathbf{H}_{s:t}](x_s) \gamma_{0:s}(dx_{0:s}) \\ &= \int \gamma_{0:s-1}(dx'_{0:s-1}) g_{s-1}(x'_{s-1}) \gamma_{0:s}(dx_{0:s}) \delta_{x_s}(dx'_s) \mathbf{Q}_{s+1:t}[\mathbf{H}_{s:t}](x_s) \mathbf{Q}_{s+1:t}[\mathbf{H}_{s:t}](x'_s) , \end{aligned}$$

which establishes the result since by definition

$$\begin{aligned} \mathcal{Q}_{e_s,t}(dx_{0:t}, dx'_{0:t}) &= \gamma_{0:s}(dx_{0:s}) \gamma_{0:s-1}(dx'_{0:s-1}) g_{s-1}(x'_{s-1}) \delta_{x_s}(dx'_s) \\ &\quad \mathbf{Q}_{s+1:t}(x_s, dx_{s+1:t}) \mathbf{Q}_{s+1:t}(x'_s, dx'_{s+1:t}) . \end{aligned}$$

If  $s = 0$ , then  $\mathbf{G}_{0,t}[h_{0:t}](x_0) = \int h_{0:t}(x_{0:t}) \mathbf{Q}_{1:t}(x_0, dx_{1:t})$  and

$$\begin{aligned} \gamma_0(\mathbf{1})\gamma_0(\mathbf{G}_{0,t}[h_{0:t}]^2) &= \gamma_0(\mathbf{G}_{0,t}[h_{0:t}]^2) = \int \mathbf{M}_0(dx_0) \mathbf{Q}_{1:t}[h_{0:t}](x_0) \mathbf{Q}_{1:t}[h_{0:t}](x_0) \\ &= \int \mathbf{M}_0(dx_0) \delta_{x_0}(dx'_0) \mathbf{Q}_{1:t}[h_{0:t}](x_0) \mathbf{Q}_{1:t}[h_{0:t}](x'_0) \\ &= \mathcal{Q}_{e_0,t}(h_{0:t}^{\otimes 2}) . \end{aligned}$$

A.4. **Proof of Theorem 4.4.** Let  $m \in \mathbb{N}^*$  and  $N \geq 2$ . Define

$$\mathcal{I}_0^m := \{k^{1:2m} \in [N]^{2m} : k^{2i-1} \neq k^{2i}, i \in [1 : m]\} , \quad (\text{A1.5})$$

$$\mathcal{I}_1^m := \{k^{1:2m} \in [N]^{2m} : k^{2i-1} = k^{2i}, i \in [1 : m]\} . \quad (\text{A1.6})$$

Define also for any  $p \in [2m]$ ,

$$\mathcal{S}_m^p := \{k^{1:2m} \in [N]^{2m} : \text{Card}(\{k^1, k^2, \dots, k^{2m-1}, k^{2m}\}) = p\} .$$

Then  $[N]^{2m} = \bigsqcup_{p=1}^{2m} \mathcal{S}_m^p$  and

$$\mathcal{I}_0^m = \bigsqcup_{p=1}^{2m} \mathcal{I}_0^m \cap \mathcal{S}_m^p = \bigsqcup_{p=2}^{2m} \mathcal{I}_0^m \cap \mathcal{S}_m^p , \quad \mathcal{I}_1^m = \bigsqcup_{p=1}^{2m} \mathcal{I}_1^m \cap \mathcal{S}_m^p = \bigsqcup_{p=1}^m \mathcal{I}_1^m \cap \mathcal{S}_m^p . \quad (\text{A1.7})$$

The first equality holds because the tuples in  $\mathcal{I}_0^m$  must contain at least two different values and the second because tuples in  $\mathcal{I}_1^m$  contain at most  $m$  different values. The proof of Theorem 4.4 is concerned with  $m = 2$  and that of Proposition A.4 with  $m \geq 2$ . Example A.1 provides the intersections for the case  $m = 2$ .

*Example A.1.* Choose  $m = 2$  and  $N \geq 4$ . Then,

$$\begin{aligned}\mathcal{I}_0^2 \cap \mathcal{S}_2^2 &= \{k^{1:4} \in [N]^4 : k^1 \neq k^2, k^3 \neq k^4, \{k^3, k^4\} = \{k^1, k^2\}\}, \\ \mathcal{I}_0^2 \cap \mathcal{S}_2^3 &= \{k^{1:4} \in [N]^4 : k^1 \neq k^2, k^3 \neq k^4, k^3 \in \{k^1, k^2\}, k^4 \notin \{k^1, k^2\}\}, \\ &\quad \sqcup \{k^{1:4} \in [N]^4 : k^1 \neq k^2, k^3 \neq k^4, k^3 \notin \{k^1, k^2\}, k^4 \in \{k^1, k^2\}\}, \\ \mathcal{I}_0^2 \cap \mathcal{S}_2^4 &= \{k^{1:4} \in [N]^4 : k^1 \neq k^2 \neq k^3 \neq k^4\},\end{aligned}$$

with  $\text{Card}(\mathcal{I}_0^2 \cap \mathcal{S}_2^2) = 2N(N-1)$ ,  $\text{Card}(\mathcal{I}_0^2 \cap \mathcal{S}_2^3) = 4N(N-1)(N-2)$  and  $\text{Card}(\mathcal{I}_0^2 \cap \mathcal{S}_2^4) = N(N-1)(N-2)(N-3)$ . As for  $\mathcal{I}_1^2$ ,

$$\begin{aligned}\mathcal{I}_1^2 \cap \mathcal{S}_2^1 &= \{k^{1:4} \in [N]^4 : k^1 = k^2 = k^3 = k^4\}, \\ \mathcal{I}_1^2 \cap \mathcal{S}_2^2 &= \{k^{1:4} \in [N]^4 : k^1 = k^2, k^3 = k^4, k^1 \neq k^3\}.\end{aligned}$$

with  $\text{Card}(\mathcal{I}_1^2 \cap \mathcal{S}_2^1) = N$  and  $\text{Card}(\mathcal{I}_1^2 \cap \mathcal{S}_2^2) = N(N-1)$ .

*Proof of Theorem 4.4.* We proceed by induction. Throughout the proof we assume that  $N \geq 4$  for the sake of simplicity. For  $t = 0$  and  $b = \mathbf{0}$ , using (4.23) and (4.20),  $\mathcal{Q}_{0,0}^{N,\text{BS}}(h) = N^{-1}(N-1)^{-1} \sum_{i,j \in [N]^2} \mathbb{1}_{i \neq j} h(\xi_0^i, \xi_0^j)$  and  $\mathcal{Q}_{0,0}(h) = M_{0,0}^{\otimes 2}(h)$ .

$$\begin{aligned}\|\mathcal{Q}_{0,0}^{N,\text{BS}}(h) - \mathcal{Q}_{0,0}(h)\|_2^2 &= \mathbb{E} \left[ \frac{1}{N^2(N-1)^2} \sum_{i,j,i',j' \in [N]^4} \mathbb{1}_{i \neq j, i' \neq j'} \{h(\xi_0^i, \xi_0^j) - \mathcal{Q}_{0,0}(h)\} \{h(\xi_0^{i'}, \xi_0^{j'}) - \mathcal{Q}_{0,0}(h)\} \right] \\ &= \frac{t_0 + \bar{t}_0}{N^2(N-1)^2},\end{aligned}$$

where

$$\begin{aligned}t_0 &= \mathbb{E} \left[ \sum_{i,j,i',j' \in \mathcal{I}_0^2 \cap \mathcal{S}_2^4} \{h(\xi_0^i, \xi_0^j) - \mathcal{Q}_{0,0}(h)\} \{h(\xi_0^{i'}, \xi_0^{j'}) - \mathcal{Q}_{0,0}(h)\} \right], \\ \bar{t}_0 &= \mathbb{E} \left[ \sum_{i,j,i',j' \in \mathcal{I}_0^2 \cap \overline{\mathcal{S}_2^4}} \{h(\xi_0^i, \xi_0^j) - \mathcal{Q}_{0,0}(h)\} \{h(\xi_0^{i'}, \xi_0^{j'}) - \mathcal{Q}_{0,0}(h)\} \right],\end{aligned}$$

where  $\mathcal{I}_0^2 \cap \mathcal{S}_2^4$  is defined in (A1.5), (A1.7) and explicited in Example A.1, and  $\mathcal{I}_0^2 \cap \overline{\mathcal{S}_2^4} = (\mathcal{I}_0^2 \cap \mathcal{S}_2^2) \sqcup (\mathcal{I}_0^2 \cap \mathcal{S}_2^3)$ . If  $(i, j, i', j') \in \mathcal{I}_0^2 \cap \mathcal{S}_2^4$ , then  $\xi_0^i, \xi_0^j, \xi_0^{i'}$  and  $\xi_0^{j'}$  are i.i.d. Therefore,

$$\mathbb{E} \left[ \{h(\xi_0^i, \xi_0^j) - \mathcal{Q}_{0,0}(h)\} \{h(\xi_0^{i'}, \xi_0^{j'}) - \mathcal{Q}_{0,0}(h)\} \right] = \mathbb{E} \left[ \{h(\xi_0^i, \xi_0^j) - \mathcal{Q}_{0,0}(h)\} \right]^2 = 0,$$

and  $t_0 = 0$ . Hence, using the fact that because  $h$  is bounded,  $\|h - \mathcal{Q}_{0,0}(h)\|_\infty \leq 2\|h\|_\infty$  and that  $\text{Card}(\mathcal{I}_0^2 \cap \overline{\mathcal{S}_2^4}) = 4N(N-1)(N-2) + 2N(N-1)$  by Example A.1, we get

$$\begin{aligned} \|\mathcal{Q}_{0,0}^{N,\text{BS}}(h) - \mathcal{Q}_{0,0}(h)\|_2^2 &\leq \frac{\bar{t}_0}{N^2(N-1)^2} \\ &\leq \frac{4|h|_\infty^2}{N^2(N-1)^2} \sum_{i,j,i',j' \in \mathcal{I}_0^2 \cap \overline{\mathcal{S}_2^4}} 1, \\ &= \frac{4|h|_\infty^2 [4N(N-1)(N-2) + 2N(N-1)]}{N^2(N-1)^2} = \mathcal{O}(N^{-1}), \end{aligned}$$

which completes the proof for  $b = 0$ . For  $b = 1$ ,

$$\mathcal{Q}_{1,0}^{N,\text{BS}}(h) - \mathcal{Q}_{1,0}(h) = N^{-1} \sum_{i=1}^N h(\xi_0^i, \xi_0^i) - \int h(x, x) M_0(dx),$$

and since  $h$  is bounded,

$$\|\mathcal{Q}_{1,0}^{N,\text{BS}}(h) - \mathcal{Q}_{1,0}(h)\|_2^2 = N^{-1} \mathbb{V}_{M_0}[h(\xi, \xi)] = \mathcal{O}(N^{-1}),$$

where  $\mathbb{V}_{M_0}$  is the variance under  $M_0$ . This completes the proof for  $t = 0$ .

Let  $t > 0$ , and assume now that (4.18) holds at time  $t-1$ . Consider the following decomposition

$$\mathcal{Q}_{b,t}^{N,\text{BS}}(h) - \mathcal{Q}_{b,t}(h) = \mathcal{Q}_{b,t}^{N,\text{BS}}(h) - \mathbb{E}[\mathcal{Q}_{b,t}^{N,\text{BS}}(h) | \mathcal{F}_{t-1}^N] + \mathbb{E}[\mathcal{Q}_{b,t}^{N,\text{BS}}(h) | \mathcal{F}_{t-1}^N] - \mathcal{Q}_{b,t}(h),$$

which, by Proposition 4.3, becomes

$$\begin{aligned} \mathcal{Q}_{b,t}^{N,\text{BS}}(h) - \mathcal{Q}_{b,t}(h) &= \mathcal{Q}_{b,t}^{N,\text{BS}}(h) - \mathcal{Q}_{b,t-1}^{N,\text{BS}}(g_{t-1}^{\otimes 2} \mathcal{M}_t^{b_t}[h]) \\ &\quad + \mathcal{Q}_{b,t}^{N,\text{BS}}(g_{t-1}^{\otimes 2} \mathcal{M}_t^{b_t}[h]) - \mathcal{Q}_{b,t}(g_{t-1}^{\otimes 2} \mathcal{M}_t^{b_t}[h]). \end{aligned} \quad (\text{A1.8})$$

By the induction hypothesis, since  $h$  is bounded and also  $g_{t-1}$  by (A2), we have

$$\lim_{N \rightarrow \infty} \left\| \mathcal{Q}_{b,t-1}^{N,\text{BS}}(g_{t-1}^{\otimes 2} \mathcal{M}_t^{b_t}[h]) - \mathcal{Q}_{b,t-1}(g_{t-1}^{\otimes 2} \mathcal{M}_t^{b_t}[h]) \right\|_2 = 0,$$

hence, by Minkowski's inequality it remains to prove that

$$\lim_{N \rightarrow \infty} \left\| \mathcal{Q}_{b,t}^{N,\text{BS}}(h) - \mathcal{Q}_{b,t-1}^{N,\text{BS}}(g_{t-1}^{\otimes 2} \mathcal{M}_t^{b_t}[h]) \right\|_2 = 0. \quad (\text{A1.9})$$

By Proposition 4.3,  $\mathbb{E}[\mathcal{Q}_{b,t}^{N,\text{BS}}(h) | \mathcal{F}_{t-1}^N] = \mathcal{Q}_{b,t}^{N,\text{BS}}(g_{t-1}^{\otimes 2} \mathcal{M}_t^{b_t}[h])$  and

$$\mathbb{E}[\mathcal{Q}_{b,t}^{N,\text{BS}}(h) \mathcal{Q}_{b,t-1}^{N,\text{BS}}(g_{t-1}^{\otimes 2} \mathcal{M}_t^{b_t}[h])] = \mathbb{E}[\mathcal{Q}_{b,t-1}^{N,\text{BS}}(g_{t-1}^{\otimes 2} \mathcal{M}_t^{b_t}[h])^2],$$

hence,

$$\left\| \mathcal{Q}_{b,t}^{N,\text{BS}}(h) - \mathcal{Q}_{b,t-1}^{N,\text{BS}}(g_{t-1}^{\otimes 2} \mathcal{M}_t^{b_t}[h]) \right\|_2^2 = \left\| \mathcal{Q}_{b,t}^{N,\text{BS}}(h) \right\|_2^2 - \left\| \mathcal{Q}_{b,t-1}^{N,\text{BS}}(g_{t-1}^{\otimes 2} \mathcal{M}_t^{b_t}[h]) \right\|_2^2.$$



Consequently, by Proposition A.8, if  $b_t = 0$ ,

$$\begin{aligned}
& \left\| \mathcal{Q}_{b,t}^{N,\text{BS}}(h) - \mathcal{Q}_{b,t-1}^{N,\text{BS}}(g_{t-1}^{\otimes 2} \mathcal{M}_t^0[h]) \right\|_2^2 \\
& \leq \left( \frac{(N-2)(N-3)}{N(N-1)} - 1 \right) \left\| \mathcal{Q}_{b,t-1}^{N,\text{BS}}(g_{t-1}^{\otimes 2} \mathcal{M}_t^0[h]) \right\|_2^2 \\
& + \frac{N-2}{N-1} G_\infty^3 |h|_\infty^2 \int \nu(dx) \mathbb{E} \left[ \frac{\Omega_{t-1}}{N} \left\{ \mathcal{Q}_{b,t-1}^{N,\text{BS}}(m_t(\cdot, x) \otimes \mathbf{1}) \mathcal{Q}_{b,t-1}^{N,\text{BS}}(\beta_t^N(x, \cdot) \otimes \mathbf{1}) \right. \right. \\
& + \mathcal{Q}_{b,t-1}^{N,\text{BS}}(\mathbf{1} \otimes m_t(\cdot, x)) \mathcal{Q}_{b,t-1}^{N,\text{BS}}(\beta_t^N(x, \cdot) \otimes \mathbf{1}) + \mathcal{Q}_{b,t-1}^{N,\text{BS}}(m_t(\cdot, x) \otimes \mathbf{1}) \\
& \quad \times \mathcal{Q}_{b,t-1}^{N,\text{BS}}(\mathbf{1} \otimes \beta_t^N(x, \cdot)) + \mathcal{Q}_{b,t-1}^{N,\text{BS}}(\mathbf{1} \otimes m_t(\cdot, x)) \mathcal{Q}_{b,t-1}^{N,\text{BS}}(\mathbf{1} \otimes \beta_t^N(x, \cdot)) \left. \right\} \\
& + G_\infty^2 |h|_\infty^2 \int \nu^{\otimes 2}(dy, dx) \mathbb{E} \left[ \frac{\Omega_{t-1}^2}{N(N-1)} \right. \\
& \quad \times \left\{ \mathcal{Q}_{b,t-1}^{N,\text{BS}}(m_t(\cdot, x) \otimes m_t(\cdot, y)) \mathcal{Q}_{b,t-1}^{N,\text{BS}}(\beta_t^N(x, \cdot) \otimes \beta_t^N(y, \cdot)) \right. \\
& \quad \left. \left. + \mathcal{Q}_{b,t-1}^{N,\text{BS}}(m_t(\cdot, x) \otimes m_t(\cdot, y)) \mathcal{Q}_{b,t-1}^{N,\text{BS}}(\beta_t^N(y, \cdot) \otimes \beta_t^N(x, \cdot)) \right\} \right],
\end{aligned} \tag{A1.10}$$

and if  $b_t = 1$ ,

$$\begin{aligned}
& \left\| \mathcal{Q}_{b,t}^{N,\text{BS}}(h) - \mathcal{Q}_{b,t-1}^{N,\text{BS}}(g_{t-1}^{\otimes 2} \mathcal{M}_t^1[h]) \right\|_2^2 \leq \left( \frac{N-1}{N} - 1 \right) \left\| \mathcal{Q}_{b,t-1}^{N,\text{BS}}(g_{t-1}^{\otimes 2} \mathcal{M}_t^1[h]) \right\|_2^2 \\
& + G_\infty^3 |h|_\infty^2 \int \mathbb{E} \left[ \frac{\Omega_{t-1}}{N} \mathcal{Q}_{b,t-1}^{N,\text{BS}}(m_t(\cdot, x) \otimes \mathbf{1}) \mathcal{Q}_{b,t-1}^{N,\text{BS}}(\beta_t^N(x, \cdot) \otimes \mathbf{1}) \right] \nu(dx). \tag{A1.11}
\end{aligned}$$

By Proposition A.4, the first term in the r.h.s. of (A1.10) and (A1.11) is  $\mathcal{O}(N^{-1})$  in both cases because  $|g_{t-1}^{\otimes 2} \mathcal{M}_t^{b_t}[h]|_\infty \leq G_\infty^2 |h|_\infty < \infty$ . We now show that the remaining terms go to zero when  $N$  goes to infinity. Define for any  $x \in \mathbf{X}$  and  $N \in \mathbb{N}^*$ ,

$$\begin{aligned}
B_N(x) &:= \frac{\Omega_{t-1}}{N} \mathcal{Q}_{b,t-1}^{N,\text{BS}}(m_t(\cdot, x) \otimes \mathbf{1}) \mathcal{Q}_{b,t-1}^{N,\text{BS}}(\mathbf{1} \otimes \beta_t^N(x, \cdot)), \\
\tilde{B}_N(x) &:= \mathcal{Q}_{b,t-1}^{N,\text{BS}}(m_t(\cdot, x) \otimes \mathbf{1}) \mathcal{Q}_{b,t-1}^{N,\text{BS}}(\mathbf{1} \otimes \mathbf{1}), \\
\tilde{B}(x) &:= \mathcal{Q}_{b,t-1}(m_t(\cdot, x) \otimes \mathbf{1}) \mathcal{Q}_{b,t-1}(\mathbf{1} \otimes \mathbf{1}).
\end{aligned} \tag{A1.12}$$

We apply Theorem C.1 with  $f_N = \mathbb{E}[B_N]$ ,  $g_N = \mathbb{E}[\tilde{B}_N]$ ,  $g = \tilde{B}$  and  $f = 0$ . To establish i), note that  $\mathbb{E}[B_N(x)] \leq G_\infty \mathbb{E}[\tilde{B}_N(x)]$  for all  $N \in \mathbb{N}^*$  and  $x \in \mathbf{X}$ , since for all  $(x, i) \in \mathbf{X} \times [N]$ ,  $\beta_t^N(x, \xi_{t-1}^i) \leq 1$  and  $N^{-1} \Omega_{t-1} \leq G_\infty$ . Then, to prove ii), for all  $(h, f) \in \mathbb{F}_b(\mathcal{X}^{\otimes t})^2$ , by the Cauchy-Schwarz inequality,

$$\begin{aligned}
& \left| \mathbb{E} \left[ \mathcal{Q}_{b,t-1}^{N,\text{BS}}(h) \mathcal{Q}_{b,t-1}^{N,\text{BS}}(f) - \mathcal{Q}_{b,t-1}(h) \mathcal{Q}_{b,t-1}(f) \right] \right| \\
& \leq \mathbb{E} \left[ \left| \left( \mathcal{Q}_{b,t-1}^{N,\text{BS}}(h) - \mathcal{Q}_{b,t-1}(h) \right) \mathcal{Q}_{b,t-1}^{N,\text{BS}}(f) \right| \right] + \mathbb{E} \left[ \left| \left( \mathcal{Q}_{b,t-1}^{N,\text{BS}}(f) - \mathcal{Q}_{b,t-1}(f) \right) \mathcal{Q}_{b,t-1}(h) \right| \right] \\
& \leq \left\| \mathcal{Q}_{b,t-1}^{N,\text{BS}}(h) - \mathcal{Q}_{b,t-1}(h) \right\|_2 \left\| \mathcal{Q}_{b,t-1}^{N,\text{BS}}(f) \right\|_2 + \left\| \mathcal{Q}_{b,t-1}^{N,\text{BS}}(f) - \mathcal{Q}_{b,t-1}(f) \right\|_2 \left\| \mathcal{Q}_{b,t-1}(h) \right\|,
\end{aligned}$$

which goes to zero by the induction hypothesis, the fact that  $\sup_{N \in \mathbb{N}} \|\mathcal{Q}_{b,t-1}^{N,BS}(f)\|_2 < \infty$  by Proposition A.4 and  $|\mathcal{Q}_{b,t-1}(h)| < \infty$ . Hence, for all  $x \in \mathbb{X}$ ,

$$\lim_{N \rightarrow \infty} g_N(x) = g(x) \quad \text{and} \quad \lim_{N \rightarrow \infty} \mathbb{E} \left[ \mathcal{Q}_{b,t-1}^{N,BS}(\mathbf{1} \otimes \mathbf{1})^2 \right] = \mathcal{Q}_{b,t-1}(\mathbf{1} \otimes \mathbf{1})^2 .$$

Added to the fact that  $\int \tilde{B}_N(x) \nu(dx) = \mathcal{Q}_{b,t-1}^{N,BS}(\mathbf{1} \otimes \mathbf{1})^2$  and  $\int \tilde{B}(x) \nu(dx) = \mathcal{Q}_{b,t-1}(\mathbf{1} \otimes \mathbf{1})^2$ , we get by applying Fubini's theorem

$$\lim_{N \rightarrow \infty} \int \mathbb{E} [\tilde{B}_N(x)] \nu(dx) = \lim_{N \rightarrow \infty} \mathbb{E} [\mathcal{Q}_{b,t-1}^{N,BS}(\mathbf{1} \otimes \mathbf{1})^2] = \mathcal{Q}_{b,t-1}(\mathbf{1} \otimes \mathbf{1})^2 = \int \tilde{B}(x) \nu(dx) .$$

Then, for iii), first we have that  $\mathbb{E}[B_N(x)^{3/2}] \leq G_\infty^{3/2} \mathbb{E}[\tilde{B}_N(x)^{3/2}]$  and

$$\sup_{N \in \mathbb{N}} \mathbb{E}[\tilde{B}_N(x)^{3/2}] \leq \sigma_+^{3/2} \sup_{N \in \mathbb{N}} \mathbb{E}[\mathcal{Q}_{b,t-1}^{N,BS}(\mathbf{1} \otimes \mathbf{1})^3] ,$$

where the r.h.s. is finite by choosing  $m = 3$  in Proposition A.4. The family of non negative random variables  $\{B_N(x)\}_{N \in \mathbb{N}}$  is then uniformly integrable for any  $x \in \mathbb{X}$ . Indeed, for any  $x \in \mathbb{X}$ ,  $\alpha \in \mathbb{R}_+^*$  and  $N \in \mathbb{N}$ ,

$$\begin{aligned} \mathbb{E}[B_N(x) \mathbb{1}_{B_N(x) \geq \alpha}] &\leq \mathbb{E}[B_N(x)^{3/2}] / \sqrt{\alpha} \\ &\leq \sigma_+^{3/2} G_\infty^{3/2} \sup_{N \in \mathbb{N}} \mathbb{E}[\mathcal{Q}_{b,t-1}^{N,BS}(\mathbf{1} \otimes \mathbf{1})^3] / \sqrt{\alpha} , \end{aligned}$$

hence  $\lim_{\alpha \rightarrow \infty} \sup_{N \in \mathbb{N}} \mathbb{E}[B_N(x) \mathbb{1}_{B_N(x) \geq \alpha}] = 0$ . On the other hand,

$$B_N(x) = \frac{\mathcal{Q}_{b,t-1}^{N,BS}(m_t(\cdot, x) \otimes \mathbf{1}) \mathcal{Q}_{b,t-1}^{N,BS}(\mathbf{1} \otimes g_{t-1} m_t(\cdot, x))}{N \phi_{t-1}^N(m_t(\cdot, x))} ,$$

and the induction hypothesis coupled with the fact that  $\phi_{t-1}^N(m_t(\cdot, x)) \xrightarrow{\mathbb{P}} \phi_{t-1}(m_t(\cdot, x))$  with  $\phi_{t-1}(m_t(\cdot, x)) > 0$  by (A2 : 3) gives

$$\begin{aligned} &\frac{\mathcal{Q}_{b,t-1}^{N,BS}(m_t(\cdot, x) \otimes \mathbf{1}) \mathcal{Q}_{b,t-1}^{N,BS}(\mathbf{1} \otimes g_{t-1} m_t(\cdot, x))}{\phi_{t-1}^N(m_t(\cdot, x))} \\ &\quad \xrightarrow{\mathbb{P}} \frac{\mathcal{Q}_{b,t-1}(m_t(\cdot, x) \otimes \mathbf{1}) \mathcal{Q}_{b,t-1}(\mathbf{1} \otimes g_{t-1} m_t(\cdot, x))}{\phi_{t-1}(m_t(\cdot, x))} . \end{aligned}$$

Hence,  $B_N(x) \xrightarrow{\mathbb{P}} 0$ , and by uniform integrability, for any  $x \in \mathbb{X}$

$$\lim_{N \rightarrow \infty} f_N(x) = \lim_{N \rightarrow \infty} \mathbb{E}[B_N(x)] = 0 .$$

Finally, by Theorem C.1 we deduce that

$$\lim_{N \rightarrow \infty} \int \mathbb{E}[B_N(x)] \nu(dx) = \int \lim_{N \rightarrow \infty} \mathbb{E}[B_N(x)] \nu(dx) = 0 . \quad (\text{A1.13})$$

The other similar terms are treated in the same way by adapting the definitions in (A1.12). As for the second integral, define for any  $(x, y) \in \mathbb{X}^2$ ,

$$R_N(x, y) := \frac{\Omega_{t-1}^2}{N(N-1)} \mathcal{Q}_{b,t-1}^{N,BS}(m_t(\cdot, x) \otimes m_t(\cdot, y)) \mathcal{Q}_{b,t-1}^{N,BS}(\beta_t^N(x, \cdot) \otimes \beta_t^N(y, \cdot)) .$$

Then, using that

$$\begin{aligned} \int \mathcal{Q}_{b,t-1}^{N,\text{BS}}(m_t(\cdot, x) \otimes m_t(\cdot, y)) \mathcal{Q}_{b,t-1}^{N,\text{BS}}(\beta_t^N(x, \cdot) \otimes \beta_t^N(y, \cdot)) \nu(dy) \\ \leq \mathcal{Q}_{b,t-1}^{N,\text{BS}}(m_t(\cdot, x) \otimes \mathbf{1}) \mathcal{Q}_{b,t-1}^{N,\text{BS}}(\beta_t^N(x, \cdot) \otimes \mathbf{1}) , \end{aligned}$$

together with Fubini's theorem we obtain

$$0 \leq \int \mathbb{E}[R_N(x, y)] \nu^{\otimes 2}(dx, dy) \leq \frac{4G_\infty}{3} \int \mathbb{E}[B_N(x)] \nu(dx) ,$$

and by (A1.13) we get that

$$\lim_{N \rightarrow \infty} \int \mathbb{E}[R_N(x, y)] \nu^{\otimes 2}(dx, dy) = 0 .$$

The remaining term goes to zero by a similar reasoning. This completes the proof of (4.18).

For the convergence rate, by the strong mixing assumption we have that

$$\beta_t^N(x, y) \leq \frac{G_\infty \sigma_+}{\sigma_- \Omega_{t-1}} \quad \forall (x, y) \in \mathbb{X}^2 , \quad (\text{A1.14})$$

and in the case  $b_t = 0$ , we have for example that

$$\begin{aligned} \int \mathcal{Q}_{b,t-1}^{N,\text{BS}}(m_t(\cdot, x) \otimes \mathbf{1}) \mathcal{Q}_{b,t-1}^{N,\text{BS}}(\beta_t^N(x, \cdot) \otimes \mathbf{1}) \nu(dx) \\ \leq \frac{G_\infty \sigma_+}{\sigma_- \Omega_{t-1}} \int \mathcal{Q}_{b,t-1}^{N,\text{BS}}(m_t(\cdot, x) \otimes \mathbf{1}) \mathcal{Q}_{b,t-1}^{N,\text{BS}}(\mathbf{1} \otimes \mathbf{1}) \nu(dx) \\ \leq \frac{G_\infty \sigma_+}{\sigma_- \Omega_{t-1}} \mathcal{Q}_{b,t-1}^{N,\text{BS}}(\mathbf{1} \otimes \mathbf{1})^2 . \end{aligned}$$

and

$$\begin{aligned} \int \mathcal{Q}_{b,t-1}^{N,\text{BS}}(m_t(\cdot, x) \otimes m_t(\cdot, y)) \mathcal{Q}_{b,t-1}^{N,\text{BS}}(\beta_t^N(x, \cdot) \otimes \beta_t^N(y, \cdot)) \nu^{\otimes 2}(dx, dy) \\ \leq \frac{G_\infty^2 \sigma_+^2}{\sigma_-^2 \Omega_{t-1}^2} \mathcal{Q}_{b,t-1}^{N,\text{BS}}(\mathbf{1} \otimes \mathbf{1})^2 . \end{aligned}$$

Thus, replacing in (i), we get

$$\begin{aligned} \|\mathcal{Q}_{b,t}^{N,\text{BS}}(h) - \mathcal{Q}_{b,t-1}^{N,\text{BS}}(g_{t-1}^{\otimes 2} \mathcal{M}_t^0[h])\|_2^2 \\ \leq \left[ \frac{(N-2)(N-3)}{N(N-1)} - 1 \right] \|\mathcal{Q}_{b,t-1}^{N,\text{BS}}(g_{t-1}^{\otimes 2} \mathcal{M}_t^{b_t}[h])\|_2^2 \\ + \frac{2\sigma_+ G_\infty^4 |h|_\infty^2}{\sigma_-} \left[ \frac{2(N-2)}{N(N-1)} + \frac{\sigma_+}{\sigma_- N(N-1)} \right] \|\mathcal{Q}_{b,t-1}^{N,\text{BS}}(\mathbf{1})\|_2^2 . \end{aligned}$$

The case  $b_t = 1$  is handled similarly using (ii) which yields

$$\begin{aligned} \|\mathcal{Q}_{b,t}^{N,\text{BS}}(h) - \mathcal{Q}_{b,t-1}^{N,\text{BS}}(g_{t-1}^{\otimes 2} \mathcal{M}_t^1[h])\|_2^2 \\ \leq \left[ \frac{N-1}{N} - 1 \right] \|\mathcal{Q}_{b,t-1}^{N,\text{BS}}(g_{t-1}^{\otimes 2} \mathcal{M}_t^1[h])\|_2^2 + \frac{\sigma_+ G_\infty^4 |h|_\infty^2}{N} \|\mathcal{Q}_{b,t-1}^{N,\text{BS}}(\mathbf{1})\|_2^2 . \end{aligned}$$

Both upper bounds are  $\mathcal{O}(N^{-1})$  by Proposition A.4. This concludes the proof.  $\square$

**A.5. Proof of Theorem 4.5.** The proof is a straightforward adaptation of the proof in [27, Theorem 1]. Let  $h \in \mathbb{F}_b(\mathcal{X})$ . By (4.28),

$$\begin{aligned} \mathcal{V}_{\gamma,t}^{N,BS}(h) &= \sum_{s=0}^t \left( \frac{N-1}{N} \right)^s \mathcal{Q}_{e_s,t}^{N,BS}(h_t^{\otimes 2}) + N \left[ \left( \frac{N-1}{N} \right)^{t+1} - 1 \right] \mathcal{Q}_{\mathbf{0},t}^{N,BS}(h_t^{\otimes 2}) \\ &\quad + \sum_{b \in \mathcal{B}_t \setminus \{\mathbf{0}, e_{0:t}\}} N \left\{ \prod_{s=0}^t \frac{1}{N^{b_s}} \left( \frac{N-1}{N} \right)^{1-b_s} \right\} \mathcal{Q}_{b,t}^{N,BS}(h_t^{\otimes 2}) \\ &\xrightarrow{\mathbb{P}} \sum_{s=0}^t \{ \mathcal{Q}_{b,t}(h_t^{\otimes 2}) - \mathcal{Q}_{\mathbf{0},t}(h_t^{\otimes 2}) \} = \mathcal{V}_{\gamma,t}^{\infty}(h), \end{aligned}$$

where we have used that for any  $b \in \mathcal{B}_t$ ,  $\mathcal{Q}_{b,t}^{N,BS}(h_t^{\otimes 2}) \xrightarrow{\mathbb{P}} \mathcal{Q}_{b,t}(h_t^{\otimes 2})$  by Theorem 4.4.

**A.6. Proof of Theorem 4.7.** Define for any  $t \in [N]$  and  $b \in \mathcal{B}_t$ ,

$$\mathcal{G}_t^N := \sigma(\mathcal{G}_{t-1}^N \cup \sigma(\{J_{k,t-1}^i\}_{(i,k) \in [N]^2}) \cup \sigma(\{A_{t-1}^i, \xi_t^i\}_{i=1}^N)), \quad (\text{A1.15})$$

with  $\mathcal{G}_0^N = \mathcal{F}_0^N$ . In the following, we write

$$\mathcal{C}_{N,b,t} := \left\{ \prod_{s=0}^t N^{b_s} \left( \frac{N}{N-1} \right)^{1-b_s} \right\} \gamma_t^N(\mathbf{1})^2 / N^2. \quad (\text{A1.16})$$

The intermediary results used in the next proof are given in Section A.10.

*Proof of Theorem 4.7.* Let  $h \in \mathbb{F}_b(\mathcal{X}^{\otimes 2})$ . We proceed again by induction. The case  $t = 0$  is a consequence of Theorem 4.4 since  $\tilde{\mathcal{Q}}_{b,0}^{N,M}(h) = \mathcal{Q}_{b,0}^{N,BS}(h)$  for any  $b \in \mathcal{B}_0$ . Let  $t > 0$ . Similarly to Theorem 4.4 we make use of the following decomposition:

$$\begin{aligned} \tilde{\mathcal{Q}}_{b,t}^{N,M}(h) - \mathcal{Q}_{b,t}(h) &= \tilde{\mathcal{Q}}_{b,t}^{N,M}(h) - \tilde{\mathcal{Q}}_{b,t-1}^{N,M}(g_{t-1}^{\otimes 2} \mathcal{M}_t^{b_t}[h]) \\ &\quad + \tilde{\mathcal{Q}}_{b,t-1}^{N,M}(g_{t-1}^{\otimes 2} \mathcal{M}_t^{b_t}[h]) - \mathcal{Q}_{b,t-1}(g_{t-1}^{\otimes 2} \mathcal{M}_t^{b_t}[h]). \end{aligned}$$

By Minkowski's inequality and the induction hypothesis, it remains to prove that

$$\lim_{N \rightarrow \infty} \left\| \tilde{\mathcal{Q}}_{b,t}^{N,M}(h) - \tilde{\mathcal{Q}}_{b,t-1}^{N,M}(g_{t-1}^{\otimes 2} \mathcal{M}_t^{b_t}[h]) \right\|_2 = 0. \quad (\text{A1.17})$$

By Lemma A.9,  $\mathbb{E}[\tilde{\mathcal{Q}}_{b,t}^{N,M}(h) | \mathcal{G}_{t-1}^N] = \tilde{\mathcal{Q}}_{b,t}^{N,M}(g_{t-1}^{\otimes 2} \mathcal{M}_t^{b_t}[h])$  and

$$\mathbb{E}[\tilde{\mathcal{Q}}_{b,t}^{N,M}(h) \tilde{\mathcal{Q}}_{b,t-1}^{N,M}(g_{t-1}^{\otimes 2} \mathcal{M}_t^{b_t}[h])] = \mathbb{E}[\tilde{\mathcal{Q}}_{b,t-1}^{N,M}(g_{t-1}^{\otimes 2} \mathcal{M}_t^{b_t}[h])^2],$$

hence,

$$\left\| \tilde{\mathcal{Q}}_{b,t}^{N,M}(h) - \tilde{\mathcal{Q}}_{b,t}^{N,M}(g_{t-1}^{\otimes 2} \mathcal{M}_t^{b_t}[h]) \right\|_2^2 = \left\| \tilde{\mathcal{Q}}_{b,t}^{N,M}(h) \right\|_2^2 - \left\| \tilde{\mathcal{Q}}_{b,t-1}^{N,M}(g_{t-1}^{\otimes 2} \mathcal{M}_t^{b_t}[h]) \right\|_2^2.$$

By Proposition A.10, if  $b_t = 0$ ,

$$\begin{aligned} \|\tilde{\mathcal{Q}}_{b,t}^{N,M}(h)\|_2^2 &= \sum_{p=2}^4 \mathbb{E} \left[ \mathcal{C}_{N,b,t}^2 \sum_{k_t^{1:4} \in \mathcal{I}_0^2 \cap \mathcal{S}_2^p} \tilde{\mathcal{T}}_t^b(k_t^1, k_t^2) \tilde{\mathcal{T}}_t^b(k_t^3, k_t^4) \right] \\ &\leq \frac{(N-2)(N-3)}{N(N-1)} \|\tilde{\mathcal{Q}}_{b,t-1}^{N,M}(g_{t-1}^{\otimes 2} \mathcal{M}_t^0[h])\|_2^2 \\ &\quad + |h|_\infty^2 \sum_{p=2}^3 \mathbb{E} \left[ \mathcal{C}_{N,b,t}^2 \sum_{k_t^{1:4} \in \mathcal{I}_0^2 \cap \mathcal{S}_2^p} \tilde{\mathcal{T}}_t^b(k_t^1, k_t^2) \tilde{\mathcal{T}}_t^b(k_t^3, k_t^4) \right], \end{aligned}$$

and

$$\begin{aligned} &\|\tilde{\mathcal{Q}}_{b,t}^{N,M}(h) - \tilde{\mathcal{Q}}_{b,t}^{N,M}(g_{t-1}^{\otimes 2} \mathcal{M}_t^{b_t}[h])\|_2^2 \\ &\leq \left( \frac{(N-2)(N-3)}{N(N-1)} - 1 \right) \|\tilde{\mathcal{Q}}_{b,t-1}^{N,M}(g_{t-1}^{\otimes 2} \mathcal{M}_t^{b_t}[h])\|_2^2 \\ &\quad + |h|_\infty^2 \sum_{p=2}^3 \mathbb{E} \left[ \mathcal{C}_{N,b,t}^2 \sum_{k_t^{1:4} \in \mathcal{I}_0^2 \cap \mathcal{S}_2^p} \tilde{\mathcal{T}}_t^b(k_t^1, k_t^2) \tilde{\mathcal{T}}_t^b(k_t^3, k_t^4) \right]. \end{aligned}$$

By Proposition A.12, (A2) and the fact that  $h$  is bounded,  $\sup_{N \in \mathbb{N}} \|\tilde{\mathcal{Q}}_{b,t-1}^{N,M}(g_{t-1}^{\otimes 2} \mathcal{M}_t^{b_t}[h])\|_2^2 < \infty$  and

$$\lim_{N \rightarrow \infty} \left( \frac{(N-2)(N-3)}{N(N-1)} - 1 \right) \|\tilde{\mathcal{Q}}_{b,t-1}^{N,M}(g_{t-1}^{\otimes 2} \mathcal{M}_t^{b_t}[h])\|_2^2 = 0,$$

and by (i) in Proposition A.11, the second term in the r.h.s. also goes to zero, which shows (A1.17) when  $b_t = 0$ . If  $b_t = 1$ ,

$$\begin{aligned} \|\mathcal{Q}_{b,t}^{N,BS}(h)\|_2^2 &= \sum_{p=1}^2 \mathbb{E} \left[ \mathcal{C}_{N,b,t}^2 \sum_{k_t^{1:4} \in \mathcal{I}_1^2 \cap \mathcal{S}_2^p} \tilde{\mathcal{T}}_t^b(k_t^1, k_t^2) \tilde{\mathcal{T}}_t^b(k_t^3, k_t^4) \right] \\ &\leq \frac{N-1}{N} \|\tilde{\mathcal{Q}}_{b,t-1}^{N,M}(g_{t-1}^{\otimes 2} \mathcal{M}_t^1[h])\|_2^2 + |h|_\infty^2 \mathbb{E} \left[ \mathcal{C}_{N,b,t}^2 \sum_{k_t^{1:4} \in \mathcal{I}_1^2 \cap \mathcal{S}_2^1} \tilde{\mathcal{T}}_t^b(k_t^1, k_t^2) \tilde{\mathcal{T}}_t^b(k_t^3, k_t^4) \right], \end{aligned}$$

and  $\|\tilde{\mathcal{Q}}_{b,t}^{N,M}(h) - \tilde{\mathcal{Q}}_{b,t}^{N,M}(g_{t-1}^{\otimes 2} \mathcal{M}_t^{b_t}[h])\|_2^2$  goes to zero similarly to the case  $b_t = 0$  and by application of Proposition A.11.

The convergence rate follows straightforwardly by Proposition A.11 since for  $p \in \{2, 3\}$

$$\mathbb{E} \left[ \mathcal{C}_{N,b,t}^2 \sum_{k_t^{1:4} \in \mathcal{I}_0^2 \cap \mathcal{S}_2^p} \tilde{\mathcal{T}}_t^b(k_t^1, k_t^2) \tilde{\mathcal{T}}_t^b(k_t^3, k_t^4) \right] = \mathcal{O}(N^{-1}),$$

and

$$\mathbb{E} \left[ \mathcal{C}_{N,b,t}^2 \sum_{k_t^{1:4} \in \mathcal{I}_1^2 \cap \mathcal{S}_2^1} \tilde{\mathcal{T}}_t^b(k_t^1, k_t^2) \tilde{\mathcal{T}}_t^b(k_t^3, k_t^4) \right] = \mathcal{O}(N^{-1}).$$

□

**A.7. Proof of Theorem 4.8.** The proof boils down to showing a *PARIS* version of the identity (4.28). Let us first prove that for all  $t \in \mathbb{N}$  and  $(k_t^1, k_t^2) \in [N]^2$ ,

$$\sum_{b \in \mathcal{B}_t} \tilde{\mathcal{T}}_t^b(k_t^1, k_t^2) = 1. \quad (\text{A1.18})$$

We proceed by induction. If  $t = 0$ ,

$$\sum_{b \in \mathcal{B}_0} \tilde{\mathcal{T}}_0^b(k_0^1, k_0^2) = \mathbb{1}_{k_0^1 \neq k_0^2} + \mathbb{1}_{k_0^1 = k_0^2} = 1.$$

Let  $t > 0$  and assume that (A1.18) holds at  $t - 1$  for all  $(k_{t-1}^1, k_{t-1}^2) \in [N]^2$ . By the induction hypothesis,

$$\begin{aligned} & \sum_{b \in \mathcal{B}_t} \tilde{\mathcal{T}}_t^b(k_t^1, k_t^2) \\ &= \sum_{b \in \mathcal{B}_{t-1}} M^{-1} \left\{ \mathbb{1}_{k_t^1 \neq k_t^2} \sum_{i=1}^M \tilde{\mathcal{T}}_{t-1}^b(J_{k_t^1, t-1}^i, J_{k_t^2, t-1}^i) + \mathbb{1}_{k_t^1 = k_t^2} \sum_{i=1}^M \sum_{n=1}^N \mathcal{W}_{t-1}^n \tilde{\mathcal{T}}_{t-1}^b(J_{k_t^1, t-1}^i, n) \right\} \\ &= M^{-1} \sum_{i=1}^M \left\{ \mathbb{1}_{k_t^1 \neq k_t^2} \sum_{b \in \mathcal{B}_{t-1}} \tilde{\mathcal{T}}_{t-1}^b(J_{k_t^1, t-1}^i, J_{k_t^2, t-1}^i) + \mathbb{1}_{k_t^1 = k_t^2} \sum_{n=1}^N \mathcal{W}_{t-1}^n \sum_{b \in \mathcal{B}_{t-1}} \tilde{\mathcal{T}}_{t-1}^b(J_{k_t^1, t-1}^i, n) \right\} \\ &= \mathbb{1}_{k_t^1 \neq k_t^2} M^{-1} \sum_{i=1}^M 1 + \mathbb{1}_{k_t^1 = k_t^2} M^{-1} \sum_{n=1}^N \sum_{i=1}^M \mathcal{W}_{t-1}^n \\ &= \mathbb{1}_{k_t^1 \neq k_t^2} + \mathbb{1}_{k_t^1 = k_t^2} = 1. \end{aligned}$$

which proves (A1.18) at time  $t$ . Consequently, we have that for all  $h \in \mathbb{F}_b(\mathcal{X})$

$$\begin{aligned} \sum_{b \in \mathcal{B}_t} \prod_{s=0}^t N^{-b_s} \left( \frac{N-1}{N} \right)^{1-b_s} \tilde{\mathcal{Q}}_{b,t}^{N,M}(h^{\otimes 2}) &= \sum_{b \in \mathcal{B}_t} \frac{\gamma_t^N(\mathbf{1})^2}{N^2} \sum_{k_t^1, k_t^2 \in [N]^2} \tilde{\mathcal{T}}_t^b(k_t^1, k_t^2) h(\xi_t^{k_t^1}) h(\xi_t^{k_t^2}) \\ &= \frac{\gamma_t^N(\mathbf{1})^2}{N^2} \sum_{k_t^1, k_t^2 \in [N]^2} h(\xi_t^{k_t^1}) h(\xi_t^{k_t^2}) \sum_{b \in \mathcal{B}_t} \tilde{\mathcal{T}}_t^b(k_t^1, k_t^2) \\ &= \gamma_t^N(\mathbf{1})^2 \eta_t^N(h)^2 = \gamma_t^N(h)^2. \end{aligned}$$

The convergence in probability is then obtained by mimicking the proof of Theorem 4.5 and using Theorem 4.7.

**A.8. Proof of Theorem 5.2.** The proof of Theorem 5.2 requires the convergence of  $\mathbf{T}_s^N[h_{0:s}](x)$  to  $\mathbf{T}_s[h_{0:s}](x)$   $\mathbb{P}$ -a.s. for any  $x \in \mathbb{X}$ .

**Proposition A.2.** For any  $s \in \mathbb{N}^*$ , any  $x \in \mathbb{X}$  and additive functional  $h_{0:s} \in \mathbb{A}_b(\mathcal{X}^{\otimes s+1})$ ,

$$\mathbf{T}_s^N[h_{0:s}](x) \xrightarrow[N \rightarrow \infty]{a.s.} \mathbf{T}_s[h_{0:s}](x). \quad (\text{A1.19})$$

*Proof.* Let  $x \in \mathbb{X}$ . Define

$$\begin{cases} a_N := \eta_{s-1}^N \left( g_{s-1} \left\{ \mathbf{T}_{s-1}^N[\tilde{h}_{0:s-1}] f_{s-1}^x + \tilde{f}_{s-1}^x \right\} \right), \\ b_N := \eta_{s-1}^N (g_{s-1} m_s(\cdot, x)), \\ b := \eta_{s-1} (g_{s-1} m_s(\cdot, x)). \end{cases}$$

where  $f_{s-1}^x : y \mapsto m_s(y, x)$  and

$$\tilde{f}_{s-1}^x : y \mapsto m_s(y, x) \left\{ \tilde{h}_{s-1}(y, x) - \mathbf{T}_s[h_{0:s}](x) \right\}.$$

Then, we have that  $a_N/b_N = \mathbf{T}_s^N[h_{0:s}](x) - \mathbf{T}_s[h_{0:s}](x)$ . By (A4),  $(f_{s-1}^x, \tilde{f}_{s-1}^x) \in \mathbb{F}_b(\mathcal{X})^2$  for any  $x \in \mathbf{X}$  and

$$\eta_{s-1} \left( g_{s-1} \{ \mathbf{T}_{s-1}[\tilde{h}_{0:s-1}] f_{s-1}^x + \tilde{f}_{s-1}^x \} \right) = 0.$$

Hence, choosing  $f_{s-1} = f_{s-1}^x$  and  $\tilde{f}_{s-1} = \tilde{f}_{s-1}^x$  in Theorem C.2 (A3.1), there exists  $(d, \tilde{d}) \in (\mathbb{R}_+^*)^2$  such that

$$\mathbb{P}(|a_N| \geq \epsilon) \leq \tilde{d} \exp(-dN\epsilon^2). \quad (\text{A1.20})$$

On the other hand, by choosing  $f_{s-1} = f_{s-1}^x$  and  $\tilde{f}_{s-1} = 0$ , there exists  $(d', \tilde{d}') \in (\mathbb{R}_+^*)^2$  such that

$$\mathbb{P}(|b_N - b| \geq \epsilon) \leq \tilde{d}' \exp(-d'N\epsilon^2).$$

Finally, since  $|a_N/b_N| \leq |\mathbf{T}_s^N[h_{0:s}](x)| + |\mathbf{T}_s[h_{0:s}](x)| \leq 2|h_{0:s}|_\infty$   $\mathbb{P}$ -a.s. and  $b > 0$  by (A2 : 3), there exist  $(c_s, \tilde{c}_s) \in (\mathbb{R}_+^*)^2$  by Lemma C.3 such that

$$\mathbb{P}(|a_N/b_N|) = \mathbb{P}(|\mathbf{T}_s^N[h_{0:s}](x) - \mathbf{T}_s[h_{0:s}](x)| \geq \epsilon) \leq \tilde{c}_s \exp(-c_s N \epsilon^2),$$

from which (A1.19) follows by applying the Borel-Cantelli Lemma.  $\square$

**Proposition A.3.** For any  $s \in \mathbb{N}^*$  and additive functional  $h_{0:s} \in \mathbb{A}_b(\mathcal{X}^{\otimes s+1})$

$$\lim_{N \rightarrow \infty} \mathbb{E} \left[ \phi_{s-1}^N \mathbf{M}_s \left[ (\mathbf{T}_s^N[h_{0:s}] - \mathbf{T}_s[h_{0:s}])^4 \right] \right] = 0. \quad (\text{A1.21})$$

*Proof.* The proof is a straightforward application of [33, Lemma 17] (which dates back to [15]). We recall it with its proof for the sake of completeness. Define for any  $x \in \mathbf{X}$

$$\begin{cases} A_N(x) &:= |\mathbf{T}_s^N[h_{0:s}](x) - \mathbf{T}_s[h_{0:s}](x)|^4 \phi_{s-1}^N(m_s(\cdot, x)), \\ \tilde{A}_N(x) &:= \phi_{s-1}^N(m_s(\cdot, x)), \\ \tilde{A}(x) &:= \phi_{s-1}(m_s(\cdot, x)). \end{cases}$$

We apply Theorem C.1 with  $f_N = \mathbb{E}[A_N]$ ,  $g_N = \mathbb{E}[\tilde{A}_N]$ ,  $f = 0$  and  $g = \tilde{A}$ .

(i) For any  $x_s \in \mathbf{X}$ ,  $|\mathbf{T}_s^N[h_{0:s}](x_s)| \leq \int |h_{0:s}|(x_{0:s}) \mathbf{T}_s^N(x_s, dx_{0:s-1}) \leq |h_{0:s}|_\infty$ , hence

$$\mathbb{E}[A_N(x)] \leq 16|h_{0:s}|_\infty^4 \mathbb{E}[\tilde{A}_N(x)]$$

(ii) We have that  $\tilde{A}_N(x) \xrightarrow[N \rightarrow \infty]{a.s.} \tilde{A}(x)$  for any  $x \in \mathbf{X}$  by (A4), (3.5) and by the dominated convergence theorem  $\lim_{N \rightarrow \infty} \mathbb{E}[\tilde{A}_N(x)] = \mathbb{E}[\tilde{A}(x)]$ . On the other hand,  $\int \mathbb{E}[\tilde{A}_N(x)] \nu(dx) = \mathbb{E}[\phi_s^N(\mathbf{1})]$ ,  $\int \tilde{A}(x) \nu(dx) = \phi_s(\mathbf{1})$  and  $\lim_{N \rightarrow \infty} \mathbb{E}[\phi_s^N(\mathbf{1})] = \phi_s(\mathbf{1})$  again by the dominated convergence theorem. Hence

$$\lim_{N \rightarrow \infty} \int \mathbb{E}[\tilde{A}_N(x)] \nu(dx) = \lim_{N \rightarrow \infty} \mathbb{E}[\phi_s^N(\mathbf{1})] = \phi_s(\mathbf{1}) = \int \tilde{A}(x) \nu(dx). \quad (\text{A1.22})$$

(iii) By Proposition A.2, (A4) and (3.5)

$$A_N(x) \xrightarrow[N \rightarrow \infty]{a.s.} 0,$$

and since  $A_N(x) \leq 16|h_{0:s}|_\infty^4 \sigma_+ \mathbb{P}$ -a.s., by the dominated convergence theorem we get  $\lim_{N \rightarrow \infty} \mathbb{E}[A_N(x)] = 0$ .

Finally, by Theorem C.1

$$\begin{aligned} \lim_{N \rightarrow \infty} \mathbb{E} \left[ \int A_N(x) \nu(dx) \right] &= \lim_{N \rightarrow \infty} \mathbb{E} [\phi_{s-1}^N M_s [(\mathbf{T}_s^N[h_{0:s}] - \mathbf{T}_s[h_{0:s}])^4]] \\ &= \int \lim_{N \rightarrow \infty} \mathbb{E}[A_N(x)] \nu(dx) = 0. \end{aligned} \quad (\text{A1.23})$$

□

*Proof of Theorem 5.2.* We write

$$\begin{aligned} \mathbf{H}_{s,t}^N &:= \mathbf{T}_s^N[h_{0:s}]c_t + \tilde{h}_{s:t}, & \mathbf{H}_{s,t} &:= \mathbf{T}_s[h_{0:s}]d_t + \tilde{h}_{s:t} \\ \mathbf{F}_{s,t}^N &:= \mathbf{T}_s^N[f_{0:s}]c_t + \tilde{f}_{s:t}, & \mathbf{F}_{s,t} &:= \mathbf{T}_s[f_{0:s}]d_t + \tilde{f}_{s:t} \end{aligned}$$

We proceed by induction on  $t \geq s$  with  $s$  fixed. By Theorem 4.4,

$$\lim_{N \rightarrow \infty} \left\| \mathcal{Q}_{e_s,s}^{N,\text{BS}}(\mathbf{H}_{s,s} \otimes \mathbf{F}_{s,s}) - \mathcal{Q}_{e_s,s}(\mathbf{H}_{s,s} \otimes \mathbf{F}_{s,s}) \right\|_2^2 = 0.$$

Hence, by the triangle inequality it suffices to show that the difference with the "idealized" estimator goes to 0, i.e.

$$\lim_{N \rightarrow \infty} \left\| \mathcal{Q}_{e_s,s}^{N,\text{BS}}(\mathbf{H}_{s,s}^N \otimes \mathbf{F}_{s,s}^N) - \mathcal{Q}_{e_s,s}^{N,\text{BS}}(\mathbf{H}_{s,s} \otimes \mathbf{F}_{s,s}) \right\|_2^2 = 0. \quad (\text{A1.24})$$

For any  $(h_{0:s}, f_{0:s}) \in \mathbb{A}_b(\mathcal{X}^{\otimes s+1})^2$  and  $(h_s, f_s) \in \mathbb{F}_b(\mathcal{X})^2$ , by (4.20)

$$\begin{aligned} &\left\| \mathcal{Q}_{e_s,s}^{N,\text{BS}}(\mathbf{H}_{s,s}^N \otimes \mathbf{F}_{s,s}^N) - \mathcal{Q}_{e_s,s}^{N,\text{BS}}(\mathbf{H}_{s,s} \otimes \mathbf{F}_{s,s}) \right\|_2 \\ &\leq N^{-2} \left\| \sum_{i,j \in [N]^2} \frac{N^{s+1} \gamma_s^N(\mathbf{1})^2}{(N-1)^s} \mathcal{T}_s^{e_s}(i,j) \left[ \mathbf{H}_{s,s}^N(\xi_s^i) \mathbf{F}_{s,s}^N(\xi_s^j) - \mathbf{H}_{s,s}(\xi_s^i) \mathbf{F}_{s,s}(\xi_s^j) \right] \right\|_2 \\ &\leq N^{-1} \sum_{i,j \in [N]^2} \left\| \frac{N^s \gamma_s^N(\mathbf{1})^2}{(N-1)^s} \mathcal{T}_s^{e_s}(i,j) \right\|_4 \left\| \mathbf{H}_{s,s}^N(\xi_s^i) \mathbf{F}_{s,s}^N(\xi_s^j) - \mathbf{H}_{s,s}(\xi_s^i) \mathbf{F}_{s,s}(\xi_s^j) \right\|_4 \end{aligned}$$

by Cauchy-Schwarz inequality. We now show that  $\sup_{N \in \mathbb{N}} \|N^s \gamma_s^N(\mathbf{1})^2 / (N-1)^s \mathcal{T}_s^{e_s}(i,j)\|_4$  is bounded for all  $(i,j) \in [N]^2$ . We first show by induction that for any  $n \in \mathbb{N}$ ,  $\mathcal{T}_n^0(i,j) \leq \mathbb{1}_{i \neq j}$ . For all  $(i,j) \in [N]^2$ ,  $\mathcal{T}_0^0(i,j) = \mathbb{1}_{i \neq j}$ , and for any  $n > 0$ , by (4.23)

$$\begin{aligned} \mathcal{T}_n^0(i,j) &= \mathbb{1}_{i \neq j} \sum_{k,\ell \in [N]^2} \beta_n^{\text{BS}}(i,k) \beta_n^{\text{BS}}(j,\ell) \mathcal{T}_{n-1}^0(k,\ell) \\ &\leq \mathbb{1}_{i \neq j} \sum_{k,\ell \in [N]^2} \beta_n^{\text{BS}}(i,k) \beta_n^{\text{BS}}(j,\ell) \mathbb{1}_{k \neq \ell} \\ &\leq \mathbb{1}_{i \neq j} \sum_{k,\ell \in [N]^2} \beta_n^{\text{BS}}(i,k) \beta_n^{\text{BS}}(j,\ell) \leq \mathbb{1}_{i \neq j}, \end{aligned}$$



where we have used the induction hypothesis in the second line. This shows the result. Next, we have that

$$\begin{aligned}\mathcal{T}_s^{e_s}(i, j) &= \mathbb{1}_{i=j} \sum_{k, \ell \in [N]^2} \beta_s^{\text{BS}}(i, k) \mathcal{W}_{s-1}^\ell \mathcal{T}_{s-1}^0(k, \ell) \\ &\leq \mathbb{1}_{i=j} \sum_{k, \ell \in [N]^2} \beta_s^{\text{BS}}(i, k) \mathcal{W}_{s-1}^\ell \mathbb{1}_{k \neq \ell} \\ &\leq \mathbb{1}_{i=j} \sum_{k, \ell \in [N]^2} \beta_s^{\text{BS}}(i, k) \mathcal{W}_{s-1}^\ell = \mathbb{1}_{i=j} .\end{aligned}$$

Hence,  $\sup_{N \in \mathbb{N}} \|N^s \gamma_s^N(\mathbf{1})^2 / (N-1)^s \mathcal{T}_s^{e_s}(i, j)\|_4 \leq (2G_\infty^2)^s \mathbb{1}_{i=j}$ . Consequently,

$$\begin{aligned}\left\| \mathcal{Q}_{e_s, s}^{N, \text{BS}}(\mathbf{H}_{s, s}^N \otimes \mathbf{F}_{s, s}^N) - \mathcal{Q}_{e_s, t}^{N, \text{BS}}(\mathbf{H}_{s, s} \otimes \mathbf{F}_{s, s}) \right\|_2 \\ \leq \frac{(2G_\infty^2)^s}{N} \sum_{i=1}^N \left\| \mathbf{H}_{s, s}^N(\xi_s^i) \mathbf{F}_{s, s}^N(\xi_s^i) - \mathbf{H}_{s, s}(\xi_s^i) \mathbf{F}_{s, s}(\xi_s^i) \right\|_4 , \quad (\text{A1.25})\end{aligned}$$

and

$$\begin{aligned}\left\| \mathbf{H}_{s, s}^N(\xi_s^i) \mathbf{F}_{s, s}^N(\xi_s^i) - \mathbf{H}_{s, s}(\xi_s^i) \mathbf{F}_{s, s}(\xi_s^i) \right\|_4 \\ \leq \left\| (\mathbf{H}_{s, s}^N(\xi_s^i) - \mathbf{H}_{s, s}(\xi_s^i)) \mathbf{F}_{s, s}^N(\xi_s^i) \right\|_4 + \left\| (\mathbf{F}_{s, s}^N(\xi_s^i) - \mathbf{F}_{s, s}(\xi_s^i)) \mathbf{H}_{s, s}^N(\xi_s^i) \right\|_4 \\ \leq C_{f, c} \left\| \mathbf{T}_s^N[h_{0:s}](\xi_s^i) - \mathbf{T}_s[h_{0:s}](\xi_s^i) \right\|_4 + C_{h, d} \left\| \mathbf{T}_s^N[f_{0:s}](\xi_s^i) - \mathbf{T}_s[f_{0:s}](\xi_s^i) \right\|_4 .\end{aligned}$$

where  $C_{f, d} := |d_s|_\infty^4 (|f_{0:s}|_\infty + |f_s|_\infty)^4$  and  $C_{h, c} := |c_s|_\infty^4 (|h_{0:s}|_\infty + |h_s|_\infty)^4$  which are finite because  $(h_{0:s}, f_{0:s}) \in \mathbb{A}_b(\mathcal{X}^{\otimes s+1})^2$ ,  $(c_s, d_s, h_s, f_s) \in \mathbb{F}_b(\mathcal{X})^4$ . We have used that

$$|\mathbf{H}_{s, s}^N(\xi_s^i)| \leq \int |c_s|_\infty |h_{0:s}|_\infty \mathbf{T}_s^N(\xi_s^i, dx_{0:s-1}) + |h_s|_\infty = |c_s|_\infty |h_{0:s}|_\infty + |h_s|_\infty = C_{h, c} .$$

For any  $h_{0:s} \in \mathbb{F}_b(\mathcal{X}^{\otimes s+1})$ ,

$$\begin{aligned}\left\| \mathbf{T}_s^N[h_{0:s}](\xi_s^i) - \mathbf{T}_s[h_{0:s}](\xi_s^i) \right\|_4^4 \\ = \mathbb{E} \left[ \mathbb{E} \left[ \left( \mathbf{T}_s^N[h_{0:s}](\xi_s^i) - \mathbf{T}_s[h_{0:s}](\xi_s^i) \right)^4 \middle| \mathcal{F}_{t-1}^N \right] \right] \\ = \mathbb{E} \left[ \sum_{j=1}^N \mathcal{W}_{s-1}^j \int (\mathbf{T}_s^N[h_{0:s}](x) - \mathbf{T}_s[h_{0:s}](x))^4 \mathbf{M}_s(\xi_{s-1}^j, dx) \right] \\ = \mathbb{E} \left[ \phi_{s-1}^N \mathbf{M}_s \left[ (\mathbf{T}_s^N[h_{0:s}] - \mathbf{T}_s[h_{0:s}])^4 \right] \right] ,\end{aligned}$$

and replacing in (A1.25) we get

$$\begin{aligned}\left\| \mathcal{Q}_{e_s, t}^{N, \text{BS}}(\mathbf{H}_{s, s}^N \otimes \mathbf{F}_{s, t}^N) - \mathcal{Q}_{e_s, s}^{N, \text{BS}}(\mathbf{H}_{s, s} \otimes \mathbf{F}_{s, s}) \right\|_2^2 \\ \leq (2G_\infty^2)^s \left\{ C_{f, d} \mathbb{E} \left[ \phi_{s-1}^N \mathbf{M}_s \left[ (\mathbf{T}_s^N[h_{0:s}] - \mathbf{T}_s[h_{0:s}])^4 \right] \right] \right. \\ \left. + C_{h, c} \mathbb{E} \left[ \phi_{s-1}^N \mathbf{M}_s \left[ (\mathbf{T}_s^N[f_{0:s}] - \mathbf{T}_s[f_{0:s}])^4 \right] \right] \right\}\end{aligned}$$

The upperbound goes to zero by Proposition A.3 and this finishes the proof of the initialization.

Let  $t > s$  and  $h_{0:s} \in \mathbb{A}_b(\mathcal{X}^{\otimes s+1})$  an additive functional. Assume that (A1.24) holds at  $t-1$  by the induction hypothesis

$$\lim_{N \rightarrow \infty} \|\mathcal{Q}_{e_s, t-1}^{N, \text{BS}}(\mathbf{Q}_t \mathbf{H}_{s,t}^N \otimes \mathbf{Q}_t \mathbf{F}_{s,t}^N) - \mathcal{Q}_{e_s, t-1}(\mathbf{Q}_t \mathbf{H}_{s,t} \otimes \mathbf{Q}_t \mathbf{F}_{s,t})\|_2 = 0,$$

where  $\mathbf{Q}_t$  is defined in (3.1) and for example

$$\mathbf{Q}_t \mathbf{H}_{s,t}^N(x_{s:t-1}) = \mathbf{T}_s^N[h_{0:s}](x_s) \mathbf{Q}_t[c_t](x_{t-1}) + \mathbf{Q}_t[\tilde{h}_{s:t}](x_{s:t-1}),$$

where  $\mathbf{Q}_t[c_t]$  and  $\mathbf{Q}_t[\tilde{h}_{s:t}]$  are bounded by (A2), and by definition of  $\mathcal{Q}_{e_s, t}$  (3.1)

$$\mathcal{Q}_{e_s, t-1}(\mathbf{Q}_t \mathbf{H}_{s,t} \otimes \mathbf{Q}_t \mathbf{F}_{s,t}) = \mathcal{Q}_{e_s, t}(\mathbf{H}_{s,t} \otimes \mathbf{F}_{s,t}).$$

Hence, to prove (5.11) it is enough to show

$$\lim_{N \rightarrow \infty} \|\mathcal{Q}_{e_s, t}^{N, \text{BS}}(\mathbf{H}_{s,t}^N \otimes \mathbf{F}_{s,t}^N) - \mathcal{Q}_{e_s, t-1}^{N, \text{BS}}(\mathbf{Q}_t \mathbf{H}_{s,t}^N \otimes \mathbf{Q}_t \mathbf{F}_{s,t}^N)\|_2 = 0. \quad (\text{A1.26})$$

Because  $\mathbf{T}_s^N[h_{0:s}]$  and  $\mathbf{T}_s^N[f_{0:s}]$  are  $\mathcal{F}_{t-1}^N$ -measurable, by Proposition 4.3

$$\mathbb{E}[\mathcal{Q}_{e_s, t}^{N, \text{BS}}(\mathbf{H}_{s,t}^N \otimes \mathbf{F}_{s,t}^N) | \mathcal{F}_{t-1}^N] = \mathcal{Q}_{e_s, t-1}^{N, \text{BS}}(\mathbf{Q}_t \mathbf{H}_{s,t}^N \otimes \mathbf{Q}_t \mathbf{F}_{s,t}^N),$$

and thus

$$\begin{aligned} & \|\mathcal{Q}_{e_s, t}^{N, \text{BS}}(\mathbf{H}_{s,t}^N \otimes \mathbf{F}_{s,t}^N) - \mathcal{Q}_{e_s, t-1}^{N, \text{BS}}(\mathbf{Q}_t \mathbf{H}_{s,t}^N \otimes \mathbf{Q}_t \mathbf{F}_{s,t}^N)\|_2 \\ &= \|\mathcal{Q}_{e_s, t}^{N, \text{BS}}(\mathbf{H}_{s,t}^N \otimes \mathbf{F}_{s,t}^N)\|_2 - \|\mathcal{Q}_{e_s, t-1}^{N, \text{BS}}(\mathbf{Q}_t \mathbf{H}_{s,t}^N \otimes \mathbf{Q}_t \mathbf{F}_{s,t}^N)\|_2. \end{aligned} \quad (\text{A1.27})$$

Now note that Proposition A.8 is still applicable with  $h = \mathbf{H}_{s,t}^N \otimes \mathbf{F}_{s,t}^N$  although there is a slight abuse because this specific  $h$  depends on the particles up to  $s-1$  through  $\mathbf{T}_s^N[h_{0:s}]$  and  $\mathbf{T}_s^N[f_{0:s}]$ . However, as they are  $\mathcal{F}_{t-1}^N$ -measurable, Proposition A.7 is still valid and hence Proposition A.8. Additionally, this specific  $h$  is bounded almost surely since for any  $(x_{s:t}, x'_{s:t}) \in (\mathcal{X}^{t-s+1})^2$

$$|\mathbf{H}_{s,t}^N(x_{s:t}) \mathbf{F}_{s,t}^N(x'_{s:t})| \leq C := (|h_{0:s}|_\infty |c_t|_\infty + |\tilde{h}_{s:t}|_\infty)(|f_{0:s}|_\infty |d_t|_\infty + |\tilde{f}_{s:t}|_\infty)$$

and hence

$$\begin{aligned} & \|\mathcal{Q}_{e_s, t}^{N, \text{BS}}(\mathbf{H}_{s,t}^N \otimes \mathbf{F}_{s,t}^N) - \mathcal{Q}_{e_s, t-1}^{N, \text{BS}}(\mathbf{Q}_t \mathbf{H}_{s,t}^N \otimes \mathbf{Q}_t \mathbf{F}_{s,t}^N)\|_2 \\ & \leq \left[ \frac{(N-2)(N-3)}{N(N-1)} - 1 \right] \|\mathcal{Q}_{b, t-1}^{N, \text{BS}}(\mathbf{Q}_t \mathbf{H}_{s,t}^N \otimes \mathbf{Q}_t \mathbf{F}_{s,t}^N)\|_2^2 \\ & \quad + \frac{N-2}{N-1} G_\infty^3 C^2 \int \nu(dx) \mathbb{E} \left[ \frac{\Omega_{t-1}}{N} \left\{ \mathcal{Q}_{b, t-1}^{N, \text{BS}}(m_t(\cdot, x) \otimes \mathbf{1}) \mathcal{Q}_{b, t-1}^{N, \text{BS}}(\beta_t^N(x, \cdot) \otimes \mathbf{1}) \right. \right. \\ & \quad + \mathcal{Q}_{b, t-1}^{N, \text{BS}}(\mathbf{1} \otimes m_t(\cdot, x)) \mathcal{Q}_{b, t-1}^{N, \text{BS}}(\beta_t^N(x, \cdot) \otimes \mathbf{1}) + \mathcal{Q}_{b, t-1}^{N, \text{BS}}(m_t(\cdot, x) \otimes \mathbf{1}) \\ & \quad \times \mathcal{Q}_{b, t-1}^{N, \text{BS}}(\mathbf{1} \otimes \beta_t^N(x, \cdot)) + \mathcal{Q}_{b, t-1}^{N, \text{BS}}(\mathbf{1} \otimes m_t(\cdot, x)) \mathcal{Q}_{b, t-1}^{N, \text{BS}}(\mathbf{1} \otimes \beta_t^N(x, \cdot)) \Big\} \\ & \quad + \int \nu^{\otimes 2}(dy, dx) \mathbb{E} \left[ \frac{G_\infty^2 C^2 \Omega_{t-1}^2}{N(N-1)} \right. \\ & \quad \times \left\{ \mathcal{Q}_{b, t-1}^{N, \text{BS}}(m_t(\cdot, x) \otimes m_t(\cdot, y)) \mathcal{Q}_{b, t-1}^{N, \text{BS}}(\beta_t^N(x, \cdot) \otimes \beta_t^N(y, \cdot)) \right. \\ & \quad \left. \left. + \mathcal{Q}_{b, t-1}^{N, \text{BS}}(m_t(\cdot, x) \otimes m_t(\cdot, y)) \mathcal{Q}_{b, t-1}^{N, \text{BS}}(\beta_t^N(y, \cdot) \otimes \beta_t^N(x, \cdot)) \right\} \right]. \end{aligned}$$

The first term in the r.h.s. goes to zero by Proposition A.4 and the fact that  $\mathbf{Q}_t \mathbf{H}_{s,t}^N \otimes \mathbf{Q}_t \mathbf{F}_{s,t}^N$  are bounded. The remaining terms are similar to those that appear in the proof of Theorem 4.4 up to some constants and thus go to zero.

For the second part, by Theorem C.2 and Borel-Cantelli Lemma,  $\phi_{0:t|t}^N(h) \xrightarrow[N \rightarrow \infty]{a.s.} \phi_{0:t|t}(h)$ . Then, by multiple applications of Theorem 5.2 and using the bilinearity of  $\mathcal{Q}_{b,t}^{N,BS}$  and  $\mathcal{Q}_{b,t}$ , for any  $s \in [0 : t]$  and bounded additive functional  $h_t$

$$\begin{aligned}
& \mathcal{Q}_{e_s,t}^{N,BS}([g_t\{\mathbf{T}_s^N[h_{0:s}] + \tilde{h}_{s:t} - \phi_{0:t|t}^N(h_t)\}]^{\otimes 2}) \\
&= \mathcal{Q}_{e_s,t}^{N,BS}([g_t\{\mathbf{T}_s^N[h_{0:s}] + \tilde{h}_{s:t}\}]^{\otimes 2}) - \phi_{0:t|t}^N(h_t) \left( \mathcal{Q}_{e_s,t}^{N,BS}([g_t\{\mathbf{T}_s^N[h_{0:s}] + \tilde{h}_{s:t}\}] \otimes \mathbf{1}) + \right. \\
&\quad \left. + \mathcal{Q}_{e_s,t}^{N,BS}(\mathbf{1} \otimes [g_t\{\mathbf{T}_s^N[h_{0:s}] + \tilde{h}_{s:t}\}]) \right) + \phi_{0:t|t}^N(h_t)^2 \mathcal{Q}_{e_s,t}^{N,BS}(\mathbf{1} \otimes \mathbf{1}) \\
&\xrightarrow{\mathbb{P}} \mathcal{Q}_{e_s,t}([g_t\{\mathbf{T}_s[h_{0:s}] + \tilde{h}_{s:t}\}]^{\otimes 2}) - \phi_{0:t|t}(h_t) \left( \mathcal{Q}_{e_s,t}([g_t\{\mathbf{T}_s[h_{0:s}] + \tilde{h}_{s:t}\}] \otimes \mathbf{1}) + \right. \\
&\quad \left. + \mathcal{Q}_{e_s,t}(\mathbf{1} \otimes [g_t\{\mathbf{T}_s[h_{0:s}] + \tilde{h}_{s:t}\}]) \right) + \phi_{0:t|t}(h_t)^2 \mathcal{Q}_{e_s,t}(\mathbf{1} \otimes \mathbf{1}) \\
&= \mathcal{Q}_{e_s,t}([g_t\{\mathbf{T}_s[h_{0:s}] + \tilde{h}_{s:t} - \phi_{0:t|t}(h_t)\}]^{\otimes 2}),
\end{aligned}$$

from which the weak consistency of  $\mathcal{V}_{0:t|t}^{N,BS}(h)$  follows.  $\square$

**A.9. Supporting results for Theorem 4.4.** In this section we prove Proposition A.4 and the upper-bound of  $\|\mathcal{Q}_{b,t}^{N,BS}(h)\|_2^2$  used in the proof of Theorem 4.4.

**Proposition A.4.** Assume that (A2) holds. For any  $t \in \mathbb{N}$ ,  $b \in \mathcal{B}_t$  and  $m \in \mathbb{N}$ ,

$$\sup_{N \in \mathbb{N}} \mathbb{E} \|\mathcal{Q}_{b,t}^{N,BS}(\mathbf{1})\|_m < \infty. \quad (\text{A1.28})$$

We preface the proof with supporting lemmata.

**Lemma A.5.** For any  $p \geq 2$ , and  $N \geq 2m$

$$\text{Card}(\mathcal{I}_0^m \cap \mathcal{S}_m^p) = \mathcal{O}(N^p) \quad \text{and} \quad \text{Card}(\mathcal{I}_1^m \cap \mathcal{S}_m^p) = \mathcal{O}(N^p).$$

*Proof.* The tuples in  $\mathcal{S}_m^p$  contain  $p$  distinct elements. These  $p$  distinct elements can be selected in  $\binom{N}{p}$  ways. For each of these tuples of size  $p$ , there are  $p^{2m}$  tuples of size  $2m$  with each element taking one of the  $p$  values. These tuples of size  $2m$  contain at most  $p$  distinct elements. Hence,

$$\text{Card}(\mathcal{I}_0^m \cap \mathcal{S}_m^p) \leq \text{Card}(\mathcal{S}_m^p) \leq \binom{N}{p} p^{2m} \leq \frac{N^p}{p!} p^{2m},$$

and similarly,

$$\text{Card}(\mathcal{I}_1^m \cap \mathcal{S}_m^p) \leq \frac{N^p}{p!} p^{2m}.$$

$\square$

**Proposition A.6.** Let  $t \in \mathbb{N}^*$ ,  $m \in \mathbb{N}^*$  and  $(k_{t-1}^1, \dots, k_{t-1}^{2m}) \in [N]^{2m}$ .

(i) If  $p \in [2 : 2m]$  and  $(k_t^1, \dots, k_t^{2m}) \in \mathcal{I}_0^m \cap \mathcal{S}_m^p$ ,

$$\mathbb{E} \left[ \prod_{j=1}^{2m} \beta_t^{\text{BS}}(k_t^j, k_{t-1}^j) \middle| \mathcal{F}_{t-1}^N \right] \leq \frac{G_\infty^p}{\Omega_{t-1}^p}.$$

(ii) If  $p \in [1 : m]$  and  $(k_t^1, \dots, k_t^{2m}) \in \mathcal{I}_1^m \cap \mathcal{S}_m^p$ ,

$$\mathbb{E} \left[ \prod_{j=1}^m \beta_t^{\text{BS}}(k_t^{2j-1}, k_{t-1}^{2j-1}) \mathcal{W}_{t-1}^{k_{t-1}^{2j}} \middle| \mathcal{F}_{t-1}^N \right] \leq \frac{G_\infty^{p+m}}{\Omega_{t-1}^{p+m}}.$$

*Proof.* Let  $p \in [2 : 2m]$ . By definition there are exactly  $p$  distinct elements in each  $\mathbf{k} := (k_t^1, \dots, k_t^{2m}) \in \mathcal{I}_0^m \cap \mathcal{S}_m^p$ . Let  $\mathbf{k}_p := \{a_1, \dots, a_p\} = \{k_t^1, \dots, k_t^{2m}\}$  the set of cardinal  $p$  containing the  $p$  distinct elements in a tuple  $\mathbf{k} \in \mathcal{S}_m^p$ . Define for any  $a_i \in \mathbf{k}_p$ ,  $V_{a_i} := \{j \in [2m] : k_t^j = a_i\}$ . Each  $V_{a_i}$  is non-empty so that it is possible to pick  $j_i \in V_{a_i}$ , and by (4.11) in Lemma 4.1 and the fact that different particles are i.i.d.

$$\begin{aligned} & \mathbb{E} \left[ \prod_{j=1}^{2m} \beta_t^{\text{BS}}(k_t^j, k_{t-1}^j) \middle| \mathcal{F}_{t-1}^N \right] \\ &= \prod_{i=1}^p \mathbb{E} \left[ \prod_{j \in V_{a_i}} \beta_t^{\text{BS}}(k_t^j, k_{t-1}^j) \middle| \mathcal{F}_{t-1}^N \right] \\ &= \prod_{i=1}^p \int \left\{ \prod_{j \in V_{a_i}} \beta_t^N(\xi_t^{k_t^j}, \xi_{t-1}^{k_{t-1}^j}) \phi_{t-1}^N \mathbb{M}_t(d\xi_t^{a_i}) \right\}, \\ &= \prod_{i=1}^p \int \left\{ \prod_{j \in V_{a_i} \setminus \{j_i\}} \beta_t^N(\xi_t^{a_i}, \xi_{t-1}^{k_{t-1}^j}) \beta_t^N(\xi_t^{a_i}, \xi_{t-1}^{k_{t-1}^{j_i}}) \phi_{t-1}^N \mathbb{M}_t(d\xi_t^{a_i}) \right\}, \\ &= \prod_{i=1}^p \int \left\{ \prod_{j \in V_{a_i} \setminus \{j_i\}} \beta_t^N(\xi_t^{a_i}, \xi_{t-1}^{k_{t-1}^j}) \mathcal{W}_{t-1}^{k_{t-1}^{j_i}} \mathbb{M}_t(\xi_{t-1}^{k_{t-1}^{j_i}}, d\xi_t^{a_i}) \right\}, \end{aligned}$$

with the convention  $\prod_{\emptyset} = 1$ . Then, since for any  $(k, \ell) \in [N]^2$ ,  $\beta_t^{\text{BS}}(\xi_t^k, \xi_{t-1}^\ell) \leq 1$ , we get

$$\mathbb{E} \left[ \prod_{j=1}^{2m} \beta_t^{\text{BS}}(k_t^j, k_{t-1}^j) \middle| \mathcal{F}_{t-1}^N \right] \leq \prod_{i=1}^p \frac{\omega_{t-1}^{k_{t-1}^{j_i}}}{\Omega_{t-1}} \int \mathbb{M}_t(\xi_{t-1}^{k_{t-1}^{j_i}}, d\xi_t^{a_i}) \leq \frac{G_\infty^p}{\Omega_{t-1}^p}.$$

Now let  $p \in [1 : m]$  and  $\mathbf{k} = (k_t^1, \dots, k_t^{2m}) \in \mathcal{I}_1^m \cap \mathcal{S}_m^p$ . Define  $\tilde{V}_{a_i} := V_{a_i} \cap \{1, 3, \dots, 2m-1\}$  for any  $a_i \in \mathbf{k}_p$ . These sets are non-empty since each  $V_{a_i}$  is non-empty and has as many even indices as odd indices, by definition of  $\mathcal{I}_1^m \cap \mathcal{S}_m^p$ . It is then possible to pick  $j_i \in \tilde{V}_{a_i}$  and, since for any  $i \in [N]$

$\mathcal{W}_{t-1}^i$  is  $\mathcal{F}_{t-1}^N$ -measurable,

$$\begin{aligned}
& \mathbb{E} \left[ \prod_{j=1}^m \beta_t^{\text{BS}}(k_t^{2j-1}, k_{t-1}^{2j-1}) \mathcal{W}_{t-1}^{k_t^{2j}} \middle| \mathcal{F}_{t-1}^N \right] \\
&= \prod_{j=1}^m \mathcal{W}_{t-1}^{k_t^{2j}} \mathbb{E} \left[ \prod_{j=1}^m \beta_t^{\text{BS}}(k_t^{2j-1}, k_{t-1}^{2j-1}) \middle| \mathcal{F}_{t-1}^N \right], \\
&= \prod_{j=1}^m \mathcal{W}_{t-1}^{k_t^{2j}} \prod_{i=1}^p \int \left\{ \prod_{j \in \tilde{V}_{a_i}} \beta_t^{\text{BS}}(k_t^j, k_{t-1}^j) \phi_{t-1}^N \mathbf{M}_t(d\xi_t^{a_i}) \right\}, \\
&= \prod_{j=1}^m \mathcal{W}_{t-1}^{k_t^{2j}} \prod_{i=1}^p \int \left\{ \prod_{j \in \tilde{V}_{a_i} \setminus \{j_i\}} \beta_t^N(\xi_t^{a_i}, \xi_{t-1}^{k_{t-1}^j}) \beta_t^N(\xi_t^{a_i}, \xi_{t-1}^{k_{t-1}^{j_i}}) \phi_{t-1}^N \mathbf{M}_t(d\xi_t^{a_i}) \right\}, \\
&= \prod_{j=1}^m \mathcal{W}_{t-1}^{k_t^{2j}} \prod_{i=1}^p \int \left\{ \prod_{j \in \tilde{V}_{a_i} \setminus \{j_i\}} \beta_t^N(\xi_t^{a_i}, \xi_{t-1}^{k_{t-1}^j}) \mathcal{W}_{t-1}^{k_{t-1}^{j_i}} \mathbf{M}_t(\xi_{t-1}^{k_{t-1}^{j_i}}, d\xi_t^{a_i}) \right\},
\end{aligned}$$

with the convention  $\prod_{\emptyset} = 1$ . Hence,

$$\mathbb{E} \left[ \prod_{j=1}^m \beta_t^{\text{BS}}(k_t^{2j-1}, k_{t-1}^{2j-1}) \mathcal{W}_{t-1}^{k_t^{2j}} \middle| \mathcal{F}_{t-1}^N \right] \leq \frac{G_{\infty}^{p+m}}{\Omega_{t-1}^{p+m}}.$$

□

*Proof of proposition A.4.* Let  $m \in \mathbb{N}$  and assume for now that  $N \geq 2m$ . We proceed by induction. For  $t = 0$ , and  $b_0 = 0$  we have that  $\mathcal{Q}_{0,0}^{N,\text{BS}}(\mathbf{1}) = N^{-1}(N-1)^{-1} \sum_{i,j \in [N]^2} \mathbb{1}_{i \neq j} = 1$  which completes the proof. If  $b_0 = 1$ ,  $\mathcal{Q}_{1,0}^{N,\text{BS}}(\mathbf{1}) = N^{-1} \sum_{i,j \in [N]^2} \mathbb{1}_{i=j} = 1$  and the result follows.

Let  $t \in \mathbb{N}^*$  and  $b \in \mathcal{B}_t$ . Assume (A1.28) holds at time  $t-1$ . Again we treat the cases  $b_t = 0$  and  $b_t = 1$  separately. In the case  $b_t = 0$ , by (A1.2),

$$\mathbb{E} \left[ \mathcal{Q}_{b,t}^{N,\text{BS}}(\mathbf{1})^m \right] = \mathbb{E} \left[ \sum_{k_{0:t}^{1:2m} \in [N]^{2m(t+1)}} \prod_{j=1}^m \bar{\Lambda}_{b,t}^{1,2}(k_{0:t}^{2j-1}, k_{0:t}^{2j}) \right]$$

and

$$\begin{aligned}
& \sum_{k_{0:t}^{1:2m} \in [N]^{2m(t+1)}} \prod_{j=1}^m \bar{\Lambda}_{b,t}^{1,2}(k_{0:t}^{2j-1}, k_{0:t}^{2j}) \\
&= \sum_{k_{0:t-1}^{1:2m} \in [N]^{2mt}} \prod_{j=1}^m \bar{\Lambda}_{b,t-1}^{1,2}(k_{0:t-1}^{2j-1}, k_{0:t-1}^{2j}) \\
&\quad \times \sum_{k_t^{1:2m} \in [N]^{2m}} \prod_{j=1}^m \frac{\Omega_{t-1}^2 \mathbb{1}_{k_t^{2j-1} \neq k_t^{2j}}}{N(N-1)} \beta_t^{\text{BS}}(k_t^{2j-1}, k_{t-1}^{2j-1}) \beta_t^{\text{BS}}(k_t^{2j}, k_{t-1}^{2j}).
\end{aligned}$$

By Proposition A.6(i) and Lemma A.5, for any  $k_{t-1}^{1:2m} \in [N]^{2m}$ ,

$$\begin{aligned} & \sum_{k_t^{1:2m} \in [N]^{2m}} \mathbb{E} \left[ \prod_{j=1}^m \beta_t^{\text{BS}}(k_t^{2j-1}, k_{t-1}^{2j-1}) \beta_t^{\text{BS}}(k_t^{2j}, k_{t-1}^{2j}) \mathbb{1}_{k_t^{2j-1} \neq k_t^{2j}} \middle| \mathcal{F}_{t-1}^N \right] \\ &= \sum_{p=2}^{2m} \sum_{k_t^{1:2m} \in \mathcal{I}_0^m \cap \mathcal{S}_m^p} \mathbb{E} \left[ \prod_{j=1}^{2m} \beta_t^{\text{BS}}(k_t^j, k_{t-1}^j) \middle| \mathcal{F}_{t-1}^N \right] \\ &\leq \sum_{p=2}^{2m} \sum_{k_t^{1:2m} \in \mathcal{I}_0^m \cap \mathcal{S}_m^p} \frac{G_\infty^p}{\Omega_{t-1}^p} \leq \sum_{p=2}^{2m} \frac{G_\infty^p \text{Card}(\mathcal{I}_0^m \cap \mathcal{S}_m^p)}{\Omega_{t-1}^p} \leq C \sum_{p=2}^{2m} \frac{N^p}{\Omega_{t-1}^p}, \end{aligned}$$

where  $C$  is a constant independent of  $N$ . Consequently, using the fact that

$$\sum_{k_{0:t-1}^{1:2m} \in [N]^{2mt}} \prod_{j=1}^m \bar{\Lambda}_{b,t-1}^{1,2}(k_{0:t-1}^{2j-1}, k_{0:t-1}^{2j}) = \mathcal{Q}_{b,t-1}^{N,\text{BS}}(\mathbf{1})^m$$

which is  $\mathcal{F}_{t-1}^N$ -measurable and that  $\Omega_{t-1} \leq NG_\infty$   $\mathbb{P}$ -a.s. by (A2), we get

$$\mathbb{E} \left[ \mathcal{Q}_{b,t}^{N,\text{BS}}(\mathbf{1})^m \right] \leq C \mathbb{E} \left[ \mathcal{Q}_{b,t-1}^{N,\text{BS}}(\mathbf{1})^m \frac{N^{2m-p} N^p}{N^m (N-1)^m} \right] \leq C \mathbb{E} \left[ \mathcal{Q}_{b,t-1}^{N,\text{BS}}(\mathbf{1})^m \right],$$

which completes the proof. In the case  $b_t = 1$ , again by (A1.2),

$$\begin{aligned} & \sum_{k_{0:t}^{1:2m} \in [N]^{2m(t+1)}} \prod_{j=1}^m \bar{\Lambda}_{b,t}^{1,2}(k_{0:t}^{2j-1}, k_{0:t}^{2j}) \\ &= \sum_{k_{0:t-1}^{1:2m} \in [N]^{2mt}} \prod_{j=1}^m \bar{\Lambda}_{b,t-1}^{1,2}(k_{0:t-1}^{2j-1}, k_{0:t-1}^{2j}) \\ &\quad \times \sum_{k_t^{1:2m} \in [N]^{2m}} \prod_{j=1}^m \frac{\Omega_{t-1}^2 \mathbb{1}_{k_t^{2j-1} = k_t^{2j}}}{N} \beta_t^{\text{BS}}(k_t^{2j-1}, k_{t-1}^{2j-1}) \mathcal{W}_{t-1}^{k_t^{2j}}. \end{aligned}$$

Then, similarly to the case  $b_t = 0$ , by Proposition A.6-(ii) and Lemma A.5, for any  $k_{t-1}^{1:2m} \in [N]^{2m}$ ,

$$\begin{aligned} & \sum_{k_t^{1:2m} \in [N]^{2m}} \mathbb{E} \left[ \prod_{j=1}^m \beta_t^{\text{BS}}(k_t^{2j-1}, k_{t-1}^{2j-1}) \mathcal{W}_{t-1}^{k_t^{2j}} \mathbb{1}_{k_t^{2j-1} = k_t^{2j}} \middle| \mathcal{F}_{t-1}^N \right] \\ &= \sum_{p=1}^m \sum_{k_t^{1:2m} \in \mathcal{I}_m^1 \cap \mathcal{S}_m^p} \mathbb{E} \left[ \prod_{j=1}^m \beta_t^{\text{BS}}(k_t^{2j-1}, k_{t-1}^{2j-1}) \mathcal{W}_{t-1}^{k_t^{2j}} \middle| \mathcal{F}_{t-1}^N \right] \\ &= \sum_{p=1}^m \sum_{k_t^{1:2m} \in \mathcal{I}_m^1 \cap \mathcal{S}_m^p} \frac{G_\infty^{m+p}}{\Omega_{t-1}^{m+p}} \leq C \sum_{p=1}^m \frac{N^p}{\Omega_{t-1}^{m+p}}, \end{aligned}$$

where  $C$  is a constant independent of  $N$ , and

$$\mathbb{E} \left[ \mathcal{Q}_{b,t}^{N,\text{BS}}(\mathbf{1})^m \right] \leq C \mathbb{E} \left[ \mathcal{Q}_{b,t-1}^{N,\text{BS}}(\mathbf{1})^m \sum_{p=1}^m \frac{\Omega_{t-1}^{2m} N^p}{N^m \Omega_{t-1}^{m+p}} \right] \leq C \mathbb{E} \left[ \mathcal{Q}_{b,t-1}^{N,\text{BS}}(\mathbf{1})^m \right],$$

which completes the proof.

If  $N < 2m$  (resp.  $N < m$ ), then a tuple in  $\mathcal{I}_0^m$  (resp.  $\mathcal{I}_1^m$ ) contains at most  $N$  different elements and the proof proceeds similarly by truncating the sums over  $p$ .  $\square$

We now give more explicit computations for the case  $m = 2$ . The sets  $\mathcal{I}_0^2 \cap \mathcal{S}_2^p$  and  $\mathcal{I}_1^2 \cap \mathcal{S}_2^p$  are detailed in Example A.1.

**Proposition A.7.** *For any  $h \in \mathbb{F}_b(\mathcal{X}^{\otimes 2(t+1)})$  and  $(k_{0:t-1}^1, \dots, k_{0:t-1}^4) \in [N]^{4t}$ ,*

$$\begin{aligned} \sum_{k_t^{1:4} \in [N]^4} \mathbb{E} \left[ h^{\otimes 2}(\xi_{0:t}^{k_{0:t}^1}, \dots, \xi_{0:t}^{k_{0:t}^4}) \prod_{j=1}^4 \beta_t^{\text{BS}}(k_t^j, k_{t-1}^j) \mathbb{1}_{k_t^1 \neq k_t^2, k_t^3 \neq k_t^4} \middle| \mathcal{F}_{t-1}^N \right] \\ \leq \frac{N(N-1)(N-2)(N-3)}{\Omega_{t-1}^4} (g_{t-1}^{\otimes 2} \mathcal{M}_t^0[h])^{\otimes 2}(\xi_{0:t-1}^{k_{0:t-1}^1}, \dots, \xi_{0:t-1}^{k_{0:t-1}^4}) \\ + \frac{N(N-1)(N-2)G_\infty^3 |h|_\infty^2}{\Omega_{t-1}^3} \vartheta_t^N(\xi_{t-1}^{k_{t-1}^{1:4}}) + \frac{N(N-1)G_\infty^2 |h|_\infty^2}{\Omega_{t-1}^2} v_t^N(\xi_{t-1}^{k_{t-1}^{1:4}}), \end{aligned}$$

and

$$\begin{aligned} \sum_{k_t^{1:4} \in [N]^4} \mathbb{E} \left[ h^{\otimes 2}(\xi_{0:t}^{k_{0:t}^1}, \dots, \xi_{0:t}^{k_{0:t}^4}) \prod_{j=1}^2 \beta_t^{\text{BS}}(k_t^{2j-1}, k_{t-1}^{2j-1}) \mathcal{W}_{t-1}^{k_{t-1}^{2j}} \mathbb{1}_{k_t^1 = k_t^2, k_t^3 = k_t^4} \middle| \mathcal{F}_{t-1}^N \right] \\ \leq \frac{N(N-1)}{\Omega_{t-1}^4} (g_{t-1}^{\otimes 2} \mathcal{M}_t^1[h])^{\otimes 2}(\xi_{0:t}^{k_{0:t}^1}, \dots, \xi_{0:t}^{k_{0:t}^4}) + \frac{NG_\infty^3}{\Omega_{t-1}^3} \text{M}_t \left[ \beta_t^N(\cdot, \xi_{t-1}^{k_{t-1}^3}) \right] (\xi_{t-1}^{k_{t-1}^1}). \end{aligned}$$

where

$$\begin{aligned} \vartheta_t^N : x^{1:4} \mapsto \text{M}_t[\beta_t^N(\cdot, x^4)](x^1) + \text{M}_t[\beta_t^N(\cdot, x^4)](x^2) + \text{M}_t[\beta_t^N(\cdot, x^3)](x^1) + \text{M}_t[\beta_t^N(\cdot, x^3)](x^2), \\ v_t^N : x^{1:4} \mapsto \text{M}_t[\beta_t^N(\cdot, x^3)](x^1) \text{M}_t[\beta_t^N(\cdot, x^4)](x^2) + \text{M}_t[\beta_t^N(\cdot, x^4)](x^1) \text{M}_t[\beta_t^N(\cdot, x^3)](x^2). \end{aligned}$$

*Proof.* Let  $(k_{0:t-1}^1, \dots, k_{0:t-1}^4) \in [N]^{4t}$ . First note that

$$\begin{aligned} \sum_{k_t^{1:4} \in [N]^4} \mathbb{E} \left[ h^{\otimes 2}(\xi_{0:t}^{k_{0:t}^1}, \dots, \xi_{0:t}^{k_{0:t}^4}) \prod_{j=1}^4 \beta_t^{\text{BS}}(k_t^j, k_{t-1}^j) \mathbb{1}_{k_t^1 \neq k_t^2, k_t^3 \neq k_t^4} \middle| \mathcal{F}_{t-1}^N \right] \\ = \sum_{p=2}^4 \sum_{k_t^{1:4} \in \mathcal{I}_0^2 \cap \mathcal{S}_2^p} \mathbb{E} \left[ h^{\otimes 2}(\xi_{0:t}^{k_{0:t}^1}, \dots, \xi_{0:t}^{k_{0:t}^4}) \prod_{j=1}^4 \beta_t^{\text{BS}}(k_t^j, k_{t-1}^j) \middle| \mathcal{F}_{t-1}^N \right]. \end{aligned}$$

We compute each term herebelow. For each  $p \in [2 : 4]$  and  $\mathbf{k} := (k_t^1, \dots, k_t^4) \in \mathcal{I}_0^2 \cap \mathcal{S}_2^p$ , let  $\mathbf{k}_p := \{a_1, \dots, a_p\} = \{k_t^1, \dots, k_t^4\}$  the set of cardinal  $p$  containing the  $p$  distinct elements in a tuple  $\mathbf{k} \in \mathcal{S}_2^p$ . Define for any  $a_i \in \mathbf{k}_p$ ,  $V_{a_i} := \{j \in [4] : k_t^j = a_i\}$ .

– Let  $(k_t^1, \dots, k_t^4) \in \mathcal{I}_0^2 \cap \mathcal{S}_2^2$ . Then,  $\mathbf{k}_2 = \{k_t^1, k_t^2\}$  and we either have  $V_{k_t^1} = \{k_t^1, k_t^3\}$  and  $V_{k_t^2} = \{k_t^2, k_t^4\}$  or  $V_{k_t^1} = \{k_t^1, k_t^4\}$  and  $V_{k_t^2} = \{k_t^2, k_t^3\}$ . Assume that  $V_{k_t^1} = \{k_t^1, k_t^3\}$  and  $V_{k_t^2} = \{k_t^2, k_t^4\}$ .

Then, by (4.11) in Lemma 4.1

$$\begin{aligned}
& \mathbb{E} \left[ \prod_{j=1}^4 \beta_t^{\text{BS}}(k_t^j, k_{t-1}^j) \middle| \mathcal{F}_{t-1}^N \right] \\
&= \mathbb{E} \left[ \beta_t^{\text{BS}}(k_t^1, k_{t-1}^1) \beta_t^{\text{BS}}(k_t^1, k_{t-1}^3) \middle| \mathcal{F}_{t-1}^N \right] \mathbb{E} \left[ \beta_t^{\text{BS}}(k_t^2, k_{t-1}^2) \beta_t^{\text{BS}}(k_t^2, k_{t-1}^4) \middle| \mathcal{F}_{t-1}^N \right] \\
&= \int \beta_t^{\text{BS}}(k_t^1, k_{t-1}^3) \mathcal{W}_{t-1}^{k_t^1} \text{M}_t(\xi_{t-1}^{k_t^1}, d\xi_t^{k_t^1}) \int \beta_t^{\text{BS}}(k_t^2, k_{t-1}^4) \mathcal{W}_{t-1}^{k_t^2} \text{M}_t(\xi_{t-1}^{k_t^2}, d\xi_t^{k_t^2}) \\
&\leq \frac{G_\infty^2}{\Omega_{t-1}^2} \text{M}_t \left[ \beta_t^N(\cdot, \xi_{t-1}^{k_t^3}) \right] (\xi_{t-1}^{k_t^1}) \text{M}_t \left[ \beta_t^N(\cdot, \xi_{t-1}^{k_t^4}) \right] (\xi_{t-1}^{k_t^2}) .
\end{aligned}$$

If  $V_{k_t^1} = \{k_t^1, k_t^4\}$  and  $V_{k_t^2} = \{k_t^2, k_t^3\}$ ,

$$\mathbb{E} \left[ \prod_{j=1}^4 \beta_t^{\text{BS}}(k_t^j, k_{t-1}^j) \middle| \mathcal{F}_{t-1}^N \right] \leq \frac{G_\infty^2}{\Omega_{t-1}^2} \text{M}_t \left[ \beta_t^N(\cdot, \xi_{t-1}^{k_t^4}) \right] (\xi_{t-1}^{k_t^1}) \text{M}_t \left[ \beta_t^N(\cdot, \xi_{t-1}^{k_t^3}) \right] (\xi_{t-1}^{k_t^2}) ,$$

and

$$\begin{aligned}
& \sum_{k_t^{1:4} \in \mathcal{I}_0^2 \cap \mathcal{S}_2^2} \mathbb{E} \left[ \prod_{j=1}^4 \beta_t^{\text{BS}}(k_t^j, k_{t-1}^j) \middle| \mathcal{F}_{t-1}^N \right] \\
&\leq \frac{N(N-1)G_\infty^2}{\Omega_{t-1}^2} \left\{ \text{M}_t \left[ \beta_t^N(\cdot, \xi_{t-1}^{k_t^3}) \right] (\xi_{t-1}^{k_t^1}) \text{M}_t \left[ \beta_t^N(\cdot, \xi_{t-1}^{k_t^4}) \right] (\xi_{t-1}^{k_t^2}) \right. \\
&\quad \left. + \text{M}_t \left[ \beta_t^N(\cdot, \xi_{t-1}^{k_t^4}) \right] (\xi_{t-1}^{k_t^1}) \text{M}_t \left[ \beta_t^N(\cdot, \xi_{t-1}^{k_t^3}) \right] (\xi_{t-1}^{k_t^2}) \right\} .
\end{aligned}$$

– Let  $(k_t^1, \dots, k_t^4) \in \mathcal{I}_0^2 \cap \mathcal{S}_2^3$ . Then, either  $\mathbf{k}_3 = \{k_t^1, k_t^2, k_t^3\}$  or  $\mathbf{k}_3 = \{k_t^1, k_t^2, k_t^4\}$ . Assume that  $\mathbf{k}_3 = \{k_t^1, k_t^2, k_t^3\}$  and  $V_{k_t^1} = \{k_t^1, k_t^4\}$ . Then,

$$\begin{aligned}
& \mathbb{E} \left[ \prod_{j=1}^4 \beta_t^{\text{BS}}(k_t^j, k_{t-1}^j) \middle| \mathcal{F}_{t-1}^N \right] = \mathcal{W}_{t-1}^{k_t^1} \mathcal{W}_{t-1}^{k_t^2} \mathcal{W}_{t-1}^{k_t^3} \text{M}_t \left[ \beta_t^N(\cdot, \xi_{t-1}^{k_t^4}) \right] (\xi_{t-1}^{k_t^1}) \\
&\leq \frac{G_\infty^3}{\Omega_{t-1}^3} \text{M}_t \left[ \beta_t^N(\cdot, \xi_{t-1}^{k_t^4}) \right] (\xi_{t-1}^{k_t^1}) .
\end{aligned}$$

Applying the same reasoning to all the combinations within  $\mathcal{I}_0^2 \cap \mathcal{S}_2^3$  we get

$$\begin{aligned}
& \sum_{k_t^{1:4} \in \mathcal{I}_0^2 \cap \mathcal{S}_2^3} \mathbb{E} \left[ \prod_{j=1}^4 \beta_t^{\text{BS}}(k_t^j, k_{t-1}^j) \middle| \mathcal{F}_{t-1}^N \right] \\
&\leq \frac{N(N-1)(N-2)G_\infty^3}{\Omega_{t-1}^3} \left\{ \text{M}_t \left[ \beta_t^N(\cdot, \xi_{t-1}^{k_t^4}) \right] (\xi_{t-1}^{k_t^1}) + \text{M}_t \left[ \beta_t^N(\cdot, \xi_{t-1}^{k_t^4}) \right] (\xi_{t-1}^{k_t^2}) \right. \\
&\quad \left. + \text{M}_t \left[ \beta_t^N(\cdot, \xi_{t-1}^{k_t^3}) \right] (\xi_{t-1}^{k_t^1}) + \text{M}_t \left[ \beta_t^N(\cdot, \xi_{t-1}^{k_t^3}) \right] (\xi_{t-1}^{k_t^2}) \right\} .
\end{aligned}$$



– Let  $(k_t^1, \dots, k_t^4) \in \mathcal{I}_0^4 \cap \mathcal{S}_2^4$ . Then,  $\mathbf{k}_4 = \{k_t^1, k_t^2, k_t^3, k_t^4\}$  and

$$\begin{aligned} & \mathbb{E} \left[ h^{\otimes 2}(\xi_{0:t}^{k_t^1}, \dots, \xi_{0:t}^{k_t^4}) \prod_{j=1}^4 \beta_t^{\text{BS}}(k_t^j, k_{t-1}^j) \middle| \mathcal{F}_{t-1}^N \right] \\ &= \int h^{\otimes 2}(\xi_{0:t}^{k_t^1}, \dots, \xi_{0:t}^{k_t^4}) \prod_{j=1}^4 \beta_t^{\text{BS}}(k_t^j, k_{t-1}^j) \phi_{t-1}^N \mathbf{M}_t(d\xi_t^{k_t^j}) \\ &= \int h^{\otimes 2}(\xi_{0:t}^{k_t^1}, \dots, \xi_{0:t}^{k_t^4}) \prod_{j=1}^4 \mathcal{W}_{t-1}^{k_{t-1}^j} \mathbf{M}_t(\xi_{t-1}^{k_{t-1}^j}, d\xi_t^{k_t^j}) . \\ &= (g_{t-1}^{\otimes 2} \mathcal{M}_t^0[h])^{\otimes 2}(\xi_{0:t-1}^{k_t^1}, \dots, \xi_{0:t-1}^{k_t^4}) / \Omega_{t-1}^4 , \end{aligned}$$

which completes the proof of the first inequality. For the second inequality, write

$$\begin{aligned} & \sum_{k_t^{1:4} \in [N]^4} \mathbb{E} \left[ h^{\otimes 2}(\xi_{0:t}^{k_t^1}, \dots, \xi_{0:t}^{k_t^4}) \prod_{j=1}^2 \beta_t^{\text{BS}}(k_t^{2j-1}, k_{t-1}^{2j-1}) \mathcal{W}_{t-1}^{k_{t-1}^{2j-1}} \mathbb{1}_{k_t^1=k_t^2, k_t^3=k_t^4} \middle| \mathcal{F}_{t-1}^N \right] \\ &= \sum_{p=1}^2 \sum_{k_t^{1:4} \in \mathcal{I}_1^2 \cap \mathcal{S}_2^2} \mathbb{E} \left[ h^{\otimes 2}(\xi_{0:t}^{k_t^1}, \dots, \xi_{0:t}^{k_t^4}) \prod_{j=1}^2 \beta_t^{\text{BS}}(k_t^{2j-1}, k_{t-1}^{2j-1}) \mathcal{W}_{t-1}^{k_{t-1}^{2j-1}} \middle| \mathcal{F}_{t-1}^N \right] , \end{aligned}$$

– Let  $(k_t^1, \dots, k_t^4) \in \mathcal{I}_1^2 \cap \mathcal{S}_2^1$ . Then  $\mathbf{k}_1 = \{k_t^1\}$  and

$$\begin{aligned} & \mathbb{E} \left[ \prod_{j=1}^2 \beta_t^{\text{BS}}(k_t^{2j-1}, k_{t-1}^{2j-1}) \mathcal{W}_{t-1}^{k_{t-1}^{2j-1}} \middle| \mathcal{F}_{t-1}^N \right] \leq \mathcal{W}_{t-1}^{k_{t-1}^1} \mathcal{W}_{t-1}^{k_{t-1}^2} \mathcal{W}_{t-1}^{k_{t-1}^4} \mathbf{M}_t[\beta_t^N(\cdot, \xi_{t-1}^{k_{t-1}^3})](\xi_{t-1}^{k_{t-1}^1}) \\ & \leq \frac{G_\infty^3}{\Omega_{t-1}^3} \mathbf{M}_t \left[ \beta_t^N(\cdot, \xi_{t-1}^{k_{t-1}^3}) \right] (\xi_{t-1}^{k_{t-1}^1}) , \end{aligned}$$

and

$$\sum_{k_t^{1:4} \in \mathcal{I}_1^2 \cap \mathcal{S}_2^2} \mathbb{E} \left[ \prod_{j=1}^2 \beta_t^{\text{BS}}(k_t^{2j-1}, k_{t-1}^{2j-1}) \mathcal{W}_{t-1}^{k_{t-1}^{2j-1}} \middle| \mathcal{F}_{t-1}^N \right] \leq \frac{NG_\infty^3}{\Omega_{t-1}^3} \mathbf{M}_t \left[ \beta_t^N(\cdot, \xi_{t-1}^{k_{t-1}^3}) \right] (\xi_{t-1}^{k_{t-1}^1}) .$$

– Let  $(k_t^1, \dots, k_t^4) \in \mathcal{I}_1^2 \cap \mathcal{S}_2^2$ . Then,  $\mathbf{k}_2 = \{k_t^1, k_t^3\}$ ,  $V_{k_t^1} = \{k_t^1, k_t^2\}$  and  $V_{k_t^3} = \{k_t^3, k_t^4\}$ . Hence,

$$\begin{aligned} & \mathbb{E} \left[ h^{\otimes 2}(\xi_{0:t}^{k_t^1}, \dots, \xi_{0:t}^{k_t^4}) \prod_{j=1}^2 \beta_t^{\text{BS}}(k_t^{2j-1}, k_{t-1}^{2j-1}) \mathcal{W}_{t-1}^{k_{t-1}^{2j-1}} \middle| \mathcal{F}_{t-1}^N \right] \\ &= (g_{t-1}^{\otimes 2} \mathcal{M}_t^1[h])^{\otimes 2}(\xi_{0:t-1}^{k_t^1}, \dots, \xi_{0:t-1}^{k_t^4}) / \Omega_{t-1}^4 , \end{aligned}$$

and

$$\begin{aligned} \sum_{k_t^{1:4} \in \mathcal{I}_1^2 \cap \mathcal{S}_2^2} \mathbb{E} \left[ h^{\otimes 2}(\xi_{0:t}^{k_{0:t}^1}, \dots, \xi_{0:t}^{k_{0:t}^4}) \prod_{j=1}^2 \beta_t^{\text{BS}}(k_t^{2j-1}, k_{t-1}^{2j-1}) \mathcal{W}_{t-1}^{k_{t-1}^{2j}} \middle| \mathcal{F}_{t-1}^N \right] \\ = \frac{N(N-1)}{\Omega_{t-1}^4} (g_{t-1}^{\otimes 2} \mathcal{M}_t^1[h])^{\otimes 2}(\xi_{0:t-1}^{k_{0:t-1}^1}, \dots, \xi_{0:t-1}^{k_{0:t-1}^4}), \end{aligned}$$

which completes the proof.  $\square$

**Proposition A.8.** *For any  $t \in \mathbb{N}$ ,  $h \in \mathbb{F}_b(\mathcal{X}^{\otimes 2(t+1)})$  and  $b \in \mathcal{B}_t$ ,*

(i) *If  $b_t = 0$ ,*

$$\begin{aligned} \|\mathcal{Q}_{b,t}^{N,\text{BS}}(h)\|_2^2 \leq \frac{(N-2)(N-3)}{N(N-1)} \|\mathcal{Q}_{b,t-1}^{N,\text{BS}}(g_{t-1}^{\otimes 2} \mathcal{M}_t^0[h])\|_2^2 + \frac{N-2}{N-1} G_\infty^3 |h|_\infty^2 \mathbb{E} \left[ \frac{\Omega_{t-1}}{N} \nu(\Theta_{b,t}^N) \right] \\ + \mathbb{E} \left[ \frac{G_\infty^2 |h|_\infty^2 \Omega_{t-1}^2}{N(N-1)} \nu^{\otimes 2}(\Upsilon_{b,t}^N) \right], \end{aligned}$$

where

$$\begin{aligned} \Theta_{b,t}^N : x \mapsto [\mathcal{Q}_{b,t-1}^{N,\text{BS}}(m_t(\cdot, x) \otimes \mathbf{1}) + \mathcal{Q}_{b,t-1}^{N,\text{BS}}(\mathbf{1} \otimes m_t(\cdot, x))] \\ \times [\mathcal{Q}_{b,t-1}^{N,\text{BS}}(\beta_t^N(x, \cdot) \otimes \mathbf{1}) + \mathcal{Q}_{b,t-1}^{N,\text{BS}}(\mathbf{1} \otimes \beta_t^N(x, \cdot))], \\ \Upsilon_{b,t}^N : (x, y) \mapsto \mathcal{Q}_{b,t-1}^{N,\text{BS}}(m_t(\cdot, x) \otimes m_t(\cdot, y)) \mathcal{Q}_{b,t-1}^{N,\text{BS}}(\beta_t^N(x, \cdot) \otimes \beta_t^N(y, \cdot)) \\ + \mathcal{Q}_{b,t-1}^{N,\text{BS}}(m_t(\cdot, x) \otimes m_t(\cdot, y)) \mathcal{Q}_{b,t-1}^{N,\text{BS}}(\beta_t^N(y, \cdot) \otimes \beta_t^N(x, \cdot)). \end{aligned}$$

(ii) *If  $b_t = 1$ ,*

$$\begin{aligned} \|\mathcal{Q}_{b,t}^{N,\text{BS}}(h)\|_2^2 \leq \frac{N-1}{N} \|\mathcal{Q}_{b,t-1}^{N,\text{BS}}(g_{t-1}^{\otimes 2} \mathcal{M}_t^1[h])\|_2^2 \\ + G_\infty^3 |h|_\infty^2 \int \mathbb{E} \left[ \frac{\Omega_{t-1}}{N} \mathcal{Q}_{b,t-1}^{N,\text{BS}}(m_t(\cdot, x) \otimes \mathbf{1}) \mathcal{Q}_{b,t-1}^{N,\text{BS}}(\beta_t^N(x, \cdot) \otimes \mathbf{1}) \right] \nu(dx). \quad (\text{A1.29}) \end{aligned}$$

*Proof.* To prove (i), if  $b_t = 0$ , by (A1.2),

$$\begin{aligned} \mathcal{Q}_{b,t}^{N,\text{BS}}(h)^2 &= \sum_{k_{0:t}^{1:4} \in [N]^{4(t+1)}} \bar{\Lambda}_{b,t}^{1,2}(k_{0:t}^1, k_{0:t}^2) \bar{\Lambda}_{b,t}^{1,2}(k_{0:t}^3, k_{0:t}^4) h^{\otimes 2}(\xi_{0:t}^{k_{0:t}^1}, \dots, \xi_{0:t}^{k_{0:t}^4}) \\ &= \sum_{k_{0:t-1}^{1:4} \in [N]^{4t}} \bar{\Lambda}_{b,t}^{1,2}(k_{0:t-1}^1, k_{0:t-1}^2) \bar{\Lambda}_{b,t}^{1,2}(k_{0:t-1}^3, k_{0:t-1}^4) \\ &\quad \times \left\{ \sum_{p=2}^4 \sum_{k_t^{1:4} \in \mathcal{I}_0^2 \cap \mathcal{S}_2^2} \frac{\Omega_{t-1}^4}{N^2(N-1)^2} \prod_{j=1}^4 \beta_t^{\text{BS}}(k_t^j, k_{t-1}^j) h^{\otimes 2}(\xi_{0:t}^{k_{0:t}^1}, \dots, \xi_{0:t}^{k_{0:t}^4}) \right\}, \end{aligned}$$

and by Proposition A.7,

$$\begin{aligned}
\|\mathcal{Q}_{b,t}^{N,\text{BS}}(h)\|_2^2 &\leq \frac{(N-2)(N-3)}{N(N-1)} \left\| \mathcal{Q}_{b,t-1}^{N,\text{BS}}(g_{t-1}^{\otimes 2} \mathcal{M}_t^0[h]) \right\|_2^2 \\
&\quad + \frac{G_\infty^3 |h|_\infty^2 (N-2)}{N(N-1)} \mathbb{E} \left[ \Omega_{t-1} \sum_{k_{0:t-1}^{1:4} \in [N]^{4t}} \bar{\Lambda}_{b,t-1}^{1,2}(k_{0:t-1}^1, k_{0:t-1}^2) \bar{\Lambda}_{b,t-1}^{1,2}(k_{0:t-1}^3, k_{0:t-1}^4) \right. \\
&\quad \times \left\{ \text{M}_t \left[ \beta_t^N(\cdot, \xi_{t-1}^{k_{t-1}^4}) \right] (\xi_{t-1}^{k_{t-1}^1}) + \text{M}_t \left[ \beta_t^N(\cdot, \xi_{t-1}^{k_{t-1}^4}) \right] (\xi_{t-1}^{k_{t-1}^2}) \right. \\
&\quad \left. \left. + \text{M}_t \left[ \beta_t^N(\cdot, \xi_{t-1}^{k_{t-1}^3}) \right] (\xi_{t-1}^{k_{t-1}^1}) + \text{M}_t \left[ \beta_t^N(\cdot, \xi_{t-1}^{k_{t-1}^3}) \right] (\xi_{t-1}^{k_{t-1}^2}) \right\} \right] \\
&\quad + \frac{G_\infty^2 |h|_\infty^2}{N(N-1)} \mathbb{E} \left[ \Omega_{t-1}^2 \sum_{k_{0:t-1}^{1:4} \in [N]^{4t}} \bar{\Lambda}_{b,t-1}^{1,2}(k_{0:t-1}^1, k_{0:t-1}^2) \bar{\Lambda}_{b,t-1}^{1,2}(k_{0:t-1}^3, k_{0:t-1}^4) \right. \\
&\quad \times \left\{ \text{M}_t \left[ \beta_t^N(\cdot, \xi_{t-1}^{k_{t-1}^3}) \right] (\xi_{t-1}^{k_{t-1}^1}) \text{M}_t \left[ \beta_t^N(\cdot, \xi_{t-1}^{k_{t-1}^4}) \right] (\xi_{t-1}^{k_{t-1}^2}) \right. \\
&\quad \left. \left. + \text{M}_t \left[ \beta_t^N(\cdot, \xi_{t-1}^{k_{t-1}^4}) \right] (\xi_{t-1}^{k_{t-1}^1}) \text{M}_t \left[ \beta_t^N(\cdot, \xi_{t-1}^{k_{t-1}^3}) \right] (\xi_{t-1}^{k_{t-1}^2}) \right\} \right] .
\end{aligned}$$

Then using (A1),

$$\begin{aligned}
&\sum_{k_{0:t-1}^{1:4} \in [N]^{4t}} \bar{\Lambda}_{b,t-1}^{1,2}(k_{0:t-1}^1, k_{0:t-1}^2) \bar{\Lambda}_{b,t-1}^{1,2}(k_{0:t-1}^3, k_{0:t-1}^4) \text{M}_t[\beta_t^N(\cdot, \xi_{t-1}^{k_{t-1}^3})](\xi_{t-1}^{k_{t-1}^1}) \\
&= \int \sum_{k_{0:t-1}^{1:4} \in [N]^{4t}} \bar{\Lambda}_{b,t-1}^{1,2}(k_{0:t-1}^1, k_{0:t-1}^2) \bar{\Lambda}_{b,t-1}^{1,2}(k_{0:t-1}^3, k_{0:t-1}^4) \beta_t^N(x, \xi_{t-1}^{k_{t-1}^3}) m_t(\xi_{t-1}^{k_{t-1}^1}, x) \nu(dx) \\
&= \int \mathcal{Q}_{b,t-1}^{N,\text{BS}}(m_t(\cdot, x) \otimes \mathbf{1}) \mathcal{Q}_{b,t-1}^{N,\text{BS}}(\beta_t^N(x, \cdot) \otimes \mathbf{1}) \nu(dx) ,
\end{aligned}$$

and

$$\begin{aligned}
&\sum_{k_{0:t-1}^{1:4} \in [N]^{4t}} \bar{\Lambda}_{b,t-1}^{1,2}(k_{0:t-1}^1, k_{0:t-1}^2) \bar{\Lambda}_{b,t-1}^{1,2}(k_{0:t-1}^3, k_{0:t-1}^4) \\
&\quad \times \text{M}_t[\beta_t^N(\cdot, \xi_{t-1}^{k_{t-1}^3})](\xi_{t-1}^{k_{t-1}^1}) \text{M}_t[\beta_t^N(\cdot, \xi_{t-1}^{k_{t-1}^4})](\xi_{t-1}^{k_{t-1}^2}) \\
&= \int \mathcal{Q}_{b,t-1}^{N,\text{BS}}(m_t(\cdot, x) \otimes m_t(\cdot, y)) \mathcal{Q}_{b,t-1}^{N,\text{BS}}(\beta_t^N(x, \cdot) \otimes \beta_t^N(y, \cdot)) \nu^{\otimes 2}(dx, dy) .
\end{aligned}$$

which completes the proof of (i) by applying the same reasoning to the remaining terms. Item (ii) is obtained in the same way.  $\square$

**A.10. Supporting results for Theorem 4.7.** In this section, we prove the analogues of Propositions 4.3-A.4-A.6 and A.7 for  $\tilde{\mathcal{Q}}_{b,t}^{N,M}$ . We remind the reader that the number of sampled indices  $M$  in the *PaRIS* estimator is *fixed* and that  $\mathcal{G}_t^N$  is defined in (A1.15). Let  $\mathcal{G}_{t-1}^N \vee \xi_t^{1:N}$  be the following  $\sigma$ -algebra:

$$\mathcal{G}_{t-1}^N \vee \xi_t^{1:N} := \sigma(\mathcal{G}_{t-1}^N \cup \sigma(\xi_t^{1:N})) . \quad (\text{A1.30})$$

**Lemma A.9.** *For any  $t \in \mathbb{N}^*$ ,  $h \in \mathbb{F}(\mathcal{X}^{\otimes 2})$  and  $b \in \mathcal{B}_t$ ,*

- (i)  $\mathbb{E}[\tilde{\mathcal{Q}}_{b,t}^{N,M}(h)|\mathcal{G}_{t-1}^N] = \tilde{\mathcal{Q}}_{b,t-1}^{N,M}(g_{t-1}^{\otimes 2}\mathcal{M}_t^{b_t}[h]),$
- (ii)  $\tilde{\mathcal{Q}}_{b,t}^{N,M}(h)$  is an unbiased estimator of  $\mathcal{Q}_{b,t}(h),$

where  $\tilde{\mathcal{Q}}_{b,t}^{N,M}(h)$  is defined in (4.29).

*Proof.* We start with the case  $b_t = 0$ . For any  $(k, \ell) \in [N]^2$ ,  $J_{k,t-1}^i$  and  $J_{\ell,t-1}^i$  are independent conditionally on  $\mathcal{G}_{t-1}^N \vee \xi_t^{1:N}$  for any  $i \in [M]$  if  $k \neq \ell$  and  $\tilde{\mathcal{T}}_{t-1}^b$  is  $\mathcal{G}_{t-1}^N$ -measurable, hence, using (4.4),

$$\begin{aligned}
& \mathbb{E} \left[ \tilde{\mathcal{T}}_t^b(k, \ell) h(\xi_t^k, \xi_t^\ell) \middle| \mathcal{G}_{t-1}^N \right] \\
&= \mathbb{E} \left[ \mathbb{1}_{k \neq \ell} h(\xi_t^k, \xi_t^\ell) M^{-1} \sum_{i=1}^M \tilde{\mathcal{T}}_{t-1}^b(J_{k,t-1}^i, J_{\ell,t-1}^i) \middle| \mathcal{G}_{t-1}^N \right] \\
&= \mathbb{E} \left[ \mathbb{1}_{k \neq \ell} h(\xi_t^k, \xi_t^\ell) M^{-1} \sum_{i=1}^M \sum_{n,m \in [N]^2} \beta_t^{\text{BS}}(k, n) \beta_t^{\text{BS}}(\ell, m) \tilde{\mathcal{T}}_{t-1}^b(n, m) \middle| \mathcal{G}_{t-1}^N \right] \\
&= \sum_{n,m \in [N]^2} \tilde{\mathcal{T}}_{t-1}^b(n, m) \mathbb{E} \left[ \mathbb{1}_{k \neq \ell} \beta_t^{\text{BS}}(k, n) \beta_t^{\text{BS}}(\ell, m) h(\xi_t^k, \xi_t^\ell) \middle| \mathcal{F}_{t-1}^N \right] \\
&= \frac{\mathbb{1}_{k \neq \ell}}{\Omega_{t-1}^2} \sum_{n,m \in [N]^2} \tilde{\mathcal{T}}_{t-1}^b(n, m) (g_{t-1}^{\otimes 2} \mathcal{M}_t^0[h]) (\xi_{t-1}^n, \xi_{t-1}^m).
\end{aligned} \tag{A1.31}$$

If  $b_t = 1$ ,

$$\begin{aligned}
& \mathbb{E} \left[ \tilde{\mathcal{T}}_t^b(k, \ell) h(\xi_t^k, \xi_t^\ell) \middle| \mathcal{G}_{t-1}^N \right] \\
&= \mathbb{E} \left[ \mathbb{1}_{k=\ell} h(\xi_t^k, \xi_t^\ell) M^{-1} \sum_{i=1}^M \sum_{m=1}^N \mathcal{W}_{t-1}^m \tilde{\mathcal{T}}_{t-1}^b(J_{k,t-1}^i, m) \middle| \mathcal{G}_{t-1}^N \right] \\
&= \mathbb{E} \left[ \mathbb{1}_{k=\ell} h(\xi_t^k, \xi_t^\ell) M^{-1} \sum_{i=1}^M \sum_{n,m \in [N]^2} \beta_t^{\text{BS}}(k, n) \mathcal{W}_{t-1}^m \tilde{\mathcal{T}}_{t-1}^b(n, m) \middle| \mathcal{G}_{t-1}^N \right] \\
&= \sum_{n,m \in [N]^2} \tilde{\mathcal{T}}_{t-1}^b(n, m) \mathbb{E} \left[ \mathbb{1}_{k=\ell} \beta_t^{\text{BS}}(k, n) \mathcal{W}_{t-1}^m h(\xi_t^k, \xi_t^\ell) \middle| \mathcal{F}_{t-1}^N \right] \\
&= \frac{\mathbb{1}_{k=\ell}}{\Omega_{t-1}^2} \sum_{n,m \in [N]^2} \tilde{\mathcal{T}}_{t-1}^b(n, m) (g_{t-1}^{\otimes 2} \mathcal{M}_t^1[h]) (\xi_{t-1}^n, \xi_{t-1}^m).
\end{aligned} \tag{A1.32}$$

Consequently, if  $b_t = 0$ , by (A1.31),

$$\begin{aligned}
& \mathbb{E}[\tilde{\mathcal{Q}}_{b,t}^{N,M}(h)|\mathcal{G}_{t-1}^N] \\
&= \prod_{s=0}^{t-1} N^{b_s} \left( \frac{N}{N-1} \right)^{1-b_s} \frac{\gamma_{t-1}^N(\mathbf{1})^2}{N^2} \sum_{k,\ell \in [N]^2} \frac{\Omega_{t-1}^2}{N(N-1)} \mathbb{E} \left[ \tilde{\mathcal{T}}_t^b(k, \ell) h(\xi_t^k, \xi_t^\ell) \middle| \mathcal{G}_{t-1}^N \right] \\
&= \prod_{s=0}^{t-1} N^{b_s} \left( \frac{N}{N-1} \right)^{1-b_s} \frac{\gamma_{t-1}^N(\mathbf{1})^2}{N^2} \sum_{n,m \in [N]^2} \tilde{\mathcal{T}}_{t-1}^b(n, m) (g_{t-1}^{\otimes 2} \mathcal{M}_t^0[h]) (\xi_{t-1}^n, \xi_{t-1}^m) \\
&= \tilde{\mathcal{Q}}_{b,t-1}^{N,M}(g_{t-1}^{\otimes 2} \mathcal{M}_t^0[h]),
\end{aligned}$$

and in a similar way,  $\mathbb{E}[\tilde{\mathcal{Q}}_{b,t}^{N,M}(h)|\mathcal{G}_{t-1}^N] = \tilde{\mathcal{Q}}_{b,t-1}^{N,M}(g_{t-1}^{\otimes 2}\mathcal{M}_t^1[h])$  by (A1.32) if  $b_t = 1$ . The second item follows straightforwardly by induction and the tower property. The induction is initialized by noting that  $\tilde{\mathcal{Q}}_{b,0}^{N,M}(h)$  is equal to  $\mathcal{Q}_{b,0}^{N,BS}(h)$  which is an unbiased estimator of  $\mathcal{Q}_{b,0}(h)$  by Proposition 4.3.  $\square$

**Proposition A.10.** *Let  $t > 0$  and  $N \geq 4$ .*

(i) *If  $b_t = 0$ ,*

$$\begin{aligned} \mathbb{E}\left[\mathcal{C}_{N,b,t}^2 \sum_{k_t^{1:4} \in \mathcal{I}_0^2 \cap \mathcal{S}_2^4} h(\xi_t^{k_t^1}, \xi_t^{k_t^2}) h(\xi_t^{k_t^3}, \xi_t^{k_t^4}) \tilde{\mathcal{T}}_t^b(k_t^1, k_t^2) \tilde{\mathcal{T}}_t^b(k_t^3, k_t^4) \middle| \mathcal{G}_{t-1}^N\right] \\ = \frac{(N-2)(N-3)}{N(N-1)} \tilde{\mathcal{Q}}_{b,t-1}^{N,M}(g_{t-1}^{\otimes 2}\mathcal{M}_t^0[h])^2, \quad (\text{A1.33}) \end{aligned}$$

$$\begin{aligned} \mathbb{E}\left[\mathcal{C}_{N,b,t}^2 \sum_{k_t^{1:4} \in \mathcal{I}_0^2 \cap \mathcal{S}_2^2} \tilde{\mathcal{T}}_t^b(k_t^1, k_t^2) \tilde{\mathcal{T}}_t^b(k_t^3, k_t^4) \middle| \mathcal{G}_{t-1}^N\right] \leq \frac{G_\infty^2(M-1)\Omega_{t-1}^2}{MN(N-1)} \nu^{\otimes 2}(\tilde{\Upsilon}_{b,t}^{N,M}) \\ + \frac{G_\infty^2\Omega_{t-1}^2}{MN(N-1)} \mathcal{C}_{N,b,t-1}^2 \sum_{k_{t-1}^{1:2} \in [N]^2} \mathsf{T}_{t-1,b}^{(1)}(k_{t-1}^1, k_{t-1}^2), \quad (\text{A1.34}) \end{aligned}$$

and

$$\begin{aligned} \mathbb{E}\left[\mathcal{C}_{N,b,t}^2 \sum_{k_t^{1:4} \in \mathcal{I}_0^2 \cap \mathcal{S}_2^3} \tilde{\mathcal{T}}_t^b(k_t^1, k_t^2) \tilde{\mathcal{T}}_t^b(k_t^3, k_t^4) \middle| \mathcal{G}_{t-1}^N\right] \leq \frac{\Omega_{t-1}G_\infty^3(M-1)(N-2)}{MN(N-1)} \nu(\tilde{\Theta}_{b,t}^{N,M}) \\ + \frac{\Omega_{t-1}G_\infty^3(N-2)}{MN(N-1)} \mathcal{C}_{N,b,t-1}^2 \sum_{k_{t-1}^{1:3} \in [N]^3} \mathsf{T}_{t-1,b}^{(2)}(k_{t-1}^{1:3}), \quad (\text{A1.35}) \end{aligned}$$

where

$$\begin{aligned} \tilde{\Theta}_{b,t}^{N,M} : x \mapsto [\tilde{\mathcal{Q}}_{b,t-1}^{N,M}(m_t(\cdot, x) \otimes \mathbf{1}) + \tilde{\mathcal{Q}}_{b,t-1}^{N,M}(\mathbf{1} \otimes m_t(\cdot, x))] \\ \times [\tilde{\mathcal{Q}}_{b,t-1}^{N,M}(\beta_t^N(x, \cdot) \otimes \mathbf{1}) + \tilde{\mathcal{Q}}_{b,t-1}^{N,M}(\mathbf{1} \otimes \beta_t^N(x, \cdot))], \\ \tilde{\Upsilon}_{b,t}^{N,M} : (x, y) \mapsto \tilde{\mathcal{Q}}_{b,t-1}^{N,M}(m_t(\cdot, x) \otimes m_t(\cdot, y)) \tilde{\mathcal{Q}}_{b,t-1}^{N,M}(\beta_t^N(x, \cdot) \otimes \beta_t^N(y, \cdot)) \\ + \tilde{\mathcal{Q}}_{b,t-1}^{N,M}(m_t(\cdot, x) \otimes m_t(\cdot, y)) \tilde{\mathcal{Q}}_{b,t-1}^{N,M}(\beta_t^N(y, \cdot) \otimes \beta_t^N(x, \cdot)), \end{aligned}$$

and

$$\begin{aligned} \mathsf{T}_{b,t}^{(1)} : (k_t^1, k_t^2) \mapsto \tilde{\mathcal{T}}_t^b(k_t^1, k_t^2)^2 + \tilde{\mathcal{T}}_t^b(k_t^1, k_t^2) \tilde{\mathcal{T}}_t^b(k_t^2, k_t^1), \\ \mathsf{T}_{b,t}^{(2)} : (k_t^1, k_t^2, k_t^3) \mapsto \tilde{\mathcal{T}}_t^b(k_t^1, k_t^2) \tilde{\mathcal{T}}_t^b(k_t^3, k_t^1) + \tilde{\mathcal{T}}_t^b(k_t^1, k_t^2) \tilde{\mathcal{T}}_t^b(k_t^1, k_t^3) \\ + \tilde{\mathcal{T}}_t^b(k_t^1, k_t^2) \tilde{\mathcal{T}}_t^b(k_t^2, k_t^3) + \tilde{\mathcal{T}}_t^b(k_t^1, k_t^2) \tilde{\mathcal{T}}_t^b(k_t^3, k_t^2). \end{aligned}$$

(ii) *If  $b_t = 1$ ,*

$$\begin{aligned} \mathbb{E}\left[\mathcal{C}_{N,b,t}^2 \sum_{k_t^{1:4} \in \mathcal{I}_1^2 \cap \mathcal{S}_2^2} h(\xi_t^{k_t^1}, \xi_t^{k_t^2}) h(\xi_t^{k_t^3}, \xi_t^{k_t^4}) \tilde{\mathcal{T}}_t^b(k_t^1, k_t^2) \tilde{\mathcal{T}}_t^b(k_t^3, k_t^4) \middle| \mathcal{G}_{t-1}^N\right] \\ = \frac{N-1}{N} \tilde{\mathcal{Q}}_{b,t-1}^{N,M}(g_{t-1}^{\otimes 2}\mathcal{M}_t^1[h])^2, \quad (\text{A1.36}) \end{aligned}$$

and

$$\begin{aligned}
& \mathbb{E} \left[ \mathcal{C}_{N,b,t}^2 \sum_{k_t^{1:4} \in \mathcal{I}_1^2 \cap \mathcal{S}_2^1} \tilde{\mathcal{T}}_t^b(k_t^1, k_t^2) \tilde{\mathcal{T}}_t^b(k_t^3, k_t^4) \middle| \mathcal{G}_{t-1}^N \right] \\
& \leq \frac{\Omega_{t-1} G_\infty^3}{N} \int \tilde{\mathcal{Q}}_{b,t-1}^{N,M}(m_t(\cdot, x) \otimes \mathbf{1}) \tilde{\mathcal{Q}}_{b,t-1}^{N,M}(\beta_t^N(x, \cdot) \otimes \mathbf{1}) \nu(dx) \\
& \quad + \frac{G_\infty^3 \Omega_{t-1}}{MN} \mathcal{C}_{N,b,t-1}^2 \sum_{k_{t-1}^{1,2,4} \in [N]^3} \tilde{\mathcal{T}}_{t-1}^b(k_{t-1}^1, k_{t-1}^2) \tilde{\mathcal{T}}_{t-1}^b(k_{t-1}^1, k_{t-1}^4) .
\end{aligned} \tag{A1.37}$$

*Proof.* We start with the case  $b_t = 0$ .

– If  $(k_t^1, \dots, k_t^4) \in \mathcal{I}_0^2 \cap \mathcal{S}_2^4$ , then  $k_t^1 \neq k_t^2 \neq k_t^3 \neq k_t^4$  and conditionally on  $\mathcal{G}_{t-1}^N \vee \xi_t^{1:N}$ ,  $J_{k_t^1, t-1}^i$ ,  $J_{k_t^2, t-1}^j$  and  $J_{k_t^4, t-1}^j$  are independent for any  $(i, j) \in [M]^2$  and  $J_{k_t^1, t-1}^{1:M} \stackrel{\text{iid}}{\sim} \beta_t^{\text{BS}}(k_t^1, \cdot)$  for any  $\ell \in [1 : 4]$ , thus, for any  $(i, j) \in [M]^2$

$$\begin{aligned}
& \mathbb{E} \left[ \tilde{\mathcal{T}}_t^b(J_{k_t^1, t-1}^i, J_{k_t^2, t-1}^j) \tilde{\mathcal{T}}_t^b(J_{k_t^3, t-1}^j, J_{k_t^4, t-1}^j) \middle| \mathcal{G}_{t-1}^N \vee \xi_t^{1:N} \right] \\
& = \mathbb{E} \left[ \tilde{\mathcal{T}}_t^b(J_{k_t^1, t-1}^i, J_{k_t^2, t-1}^j) \middle| \mathcal{G}_{t-1}^N \vee \xi_t^{1:N} \right] \mathbb{E} \left[ \tilde{\mathcal{T}}_t^b(J_{k_t^3, t-1}^j, J_{k_t^4, t-1}^j) \middle| \mathcal{G}_{t-1}^N \vee \xi_t^{1:N} \right] \\
& = \sum_{k_{t-1}^{1:4} \in [N]^4} \prod_{n=1}^4 \beta_t^{\text{BS}}(k_t^n, k_{t-1}^n) \tilde{\mathcal{T}}_{t-1}^b(k_{t-1}^1, k_{t-1}^2) \tilde{\mathcal{T}}_{t-1}^b(k_{t-1}^3, k_{t-1}^4) ,
\end{aligned}$$

and

$$\begin{aligned}
& \mathbb{E} \left[ h(\xi_t^{k_t^1}, \xi_t^{k_t^2}) h(\xi_t^{k_t^3}, \xi_t^{k_t^4}) M^{-2} \sum_{i,j \in [M]^2} \tilde{\mathcal{T}}_t^b(J_{k_t^1, t-1}^i, J_{k_t^2, t-1}^j) \tilde{\mathcal{T}}_t^b(J_{k_t^3, t-1}^j, J_{k_t^4, t-1}^j) \middle| \mathcal{G}_{t-1}^N \right] \\
& = \mathbb{E} \left[ h(\xi_t^{k_t^1}, \xi_t^{k_t^2}) h(\xi_t^{k_t^3}, \xi_t^{k_t^4}) M^{-2} \sum_{i,j \in [M]^2} \sum_{k_{t-1}^{1:4} \in [N]^4} \prod_{n=1}^4 \beta_t^{\text{BS}}(k_t^n, k_{t-1}^n) \right. \\
& \quad \left. \times \tilde{\mathcal{T}}_{t-1}^b(k_{t-1}^1, k_{t-1}^2) \tilde{\mathcal{T}}_{t-1}^b(k_{t-1}^3, k_{t-1}^4) \middle| \mathcal{G}_{t-1}^N \right] \\
& = \sum_{k_{t-1}^{1:4} \in [N]^4} \tilde{\mathcal{T}}_{t-1}^b(k_{t-1}^1, k_{t-1}^2) \tilde{\mathcal{T}}_{t-1}^b(k_{t-1}^3, k_{t-1}^4) \mathbb{E} \left[ \prod_{n=1}^4 \beta_t^{\text{BS}}(k_t^n, k_{t-1}^n) h(\xi_t^{k_t^1}, \xi_t^{k_t^2}) h(\xi_t^{k_t^3}, \xi_t^{k_t^4}) \middle| \mathcal{G}_{t-1}^N \right] \\
& = \frac{1}{\Omega_{t-1}^4} \sum_{k_{t-1}^{1:4} \in [N]^4} \tilde{\mathcal{T}}_{t-1}^b(k_{t-1}^1, k_{t-1}^2) \tilde{\mathcal{T}}_{t-1}^b(k_{t-1}^3, k_{t-1}^4) (g_{t-1}^{\otimes 2} \mathcal{M}_t^0[h])^{\otimes 2}(\xi_{t-1}^{k_{t-1}^1}, \xi_{t-1}^{k_{t-1}^2}, \xi_{t-1}^{k_{t-1}^3}, \xi_{t-1}^{k_{t-1}^4}) ,
\end{aligned}$$

where we have used that  $\tilde{\mathcal{T}}_{t-1}^b(k_{t-1}^1, k_{t-1}^2)$  and  $\tilde{\mathcal{T}}_{t-1}^b(k_{t-1}^3, k_{t-1}^4)$  are  $\mathcal{G}_{t-1}^N$ -measurable in the third equality and then proceeded similarly to Proposition A.7. By Example A.1,  $\text{Card}(\mathcal{I}_0^2 \cap \mathcal{S}_2^4) = N(N-1)(N-2)(N-3)$  and since  $b_t = 0$  implies that

$$\mathcal{C}_{N,b,t}^2 = \mathcal{C}_{N,b,t-1}^2 \frac{\Omega_{t-1}^4}{N^2(N-1)^2} , \tag{A1.38}$$

we obtain

$$\begin{aligned}
& \mathbb{E} \left[ \mathcal{C}_{N,b,t}^2 \sum_{k_t^{1:4} \in \mathcal{I}_0^2 \cap \mathcal{S}_2^4} h(\xi_t^{k_t^1}, \xi_t^{k_t^2}) h(\xi_t^{k_t^3}, \xi_t^{k_t^4}) \tilde{\mathcal{T}}_t^b(k_t^1, k_t^2) \tilde{\mathcal{T}}_t^b(k_t^3, k_t^4) \middle| \mathcal{G}_{t-1}^N \right] \\
&= \mathcal{C}_{N,b,t}^2 \sum_{k_t^{1:4} \in \mathcal{I}_0^2 \cap \mathcal{S}_2^4} \mathbb{E} \left[ h(\xi_t^{k_t^1}, \xi_t^{k_t^2}) h(\xi_t^{k_t^3}, \xi_t^{k_t^4}) \right. \\
&\quad \times M^{-2} \sum_{i,j \in [M]^2} \tilde{\mathcal{T}}_t^b(J_{k_t^1,t-1}^i, J_{k_t^2,t-1}^i) \tilde{\mathcal{T}}_t^b(J_{k_t^3,t-1}^j, J_{k_t^4,t-1}^j) \middle| \mathcal{G}_{t-1}^N \Big] \\
&= \frac{(N-2)(N-3)}{N(N-1)} \tilde{\mathcal{Q}}_{b,t-1}^{N,M} (g_{t-1}^{\otimes 2} \mathcal{M}_t^0[h])^2.
\end{aligned}$$

– If  $(k_t^1, \dots, k_t^4) \in \mathcal{I}_0^2 \cap \mathcal{S}_2^4$ . Then  $k_2 = \{k_t^1, k_t^2\}$  and we either have  $V_{k_t^1} = \{k_t^1, k_t^3\}$  and  $V_{k_t^2} = \{k_t^2, k_t^4\}$  or  $V_{k_t^1} = \{k_t^1, k_t^4\}$  and  $V_{k_t^2} = \{k_t^2, k_t^3\}$ . Assume that  $V_{k_t^1} = \{k_t^1, k_t^3\}$  and  $V_{k_t^2} = \{k_t^2, k_t^4\}$ . Taking into account that  $J_{k_t^1,t-1}^i = J_{k_t^3,t-1}^i$  and  $J_{k_t^2,t-1}^i = J_{k_t^4,t-1}^i$ , we get

$$\begin{aligned}
& \mathbb{E} \left[ \tilde{\mathcal{T}}_t^b(k_t^1, k_t^2) \tilde{\mathcal{T}}_t^b(k_t^3, k_t^4) \middle| \mathcal{G}_{t-1}^N \right] \\
&= \mathbb{E} \left[ M^{-2} \sum_{i,j \in [M]^2} \tilde{\mathcal{T}}_{t-1}^b(J_{k_t^1,t-1}^i, J_{k_t^2,t-1}^i) \tilde{\mathcal{T}}_{t-1}^b(J_{k_t^3,t-1}^j, J_{k_t^4,t-1}^j) \middle| \mathcal{G}_{t-1}^N \right] \\
&= \mathbb{E} \left[ M^{-2} \sum_{i,j \in [M]^2} \mathbb{1}_{i=j} \sum_{k_{t-1}^{1:2} \in [N]^2} \beta_t^{\text{BS}}(k_t^1, k_{t-1}^1) \beta_t^{\text{BS}}(k_t^2, k_{t-1}^2) \tilde{\mathcal{T}}_{t-1}^b(k_{t-1}^1, k_{t-1}^2)^2 \right. \\
&\quad \left. + M^{-2} \sum_{i,j \in [M]^2} \mathbb{1}_{i \neq j} \sum_{k_{t-1}^{1:4} \in [N]^4} \prod_{\ell=1}^4 \beta_t^{\text{BS}}(k_t^\ell, k_{t-1}^\ell) \tilde{\mathcal{T}}_{t-1}^b(k_{t-1}^1, k_{t-1}^2) \tilde{\mathcal{T}}_{t-1}^b(k_{t-1}^3, k_{t-1}^4) \middle| \mathcal{G}_{t-1}^N \right].
\end{aligned}$$

Then,

$$\begin{aligned}
& \mathbb{E} \left[ \tilde{\mathcal{T}}_t^b(k_t^1, k_t^2) \tilde{\mathcal{T}}_t^b(k_t^3, k_t^4) \middle| \mathcal{G}_{t-1}^N \right] \leq \frac{G_\infty^2}{M\Omega_{t-1}^2} \sum_{k_{t-1}^{1:2} \in [N]^2} \tilde{\mathcal{T}}_{t-1}^b(k_{t-1}^1, k_{t-1}^2)^2 + \frac{(M-1)G_\infty^2}{M\Omega_{t-1}^2} \\
&\times \sum_{k_{t-1}^{1:4} \in [N]^4} \text{M}_t[\beta_t^N(., \xi_{t-1}^{k_{t-1}^3})](\xi_{t-1}^{k_{t-1}^1}) \text{M}_t[\beta_t^N(., \xi_{t-1}^{k_{t-1}^4})](\xi_{t-1}^{k_{t-1}^2}) \tilde{\mathcal{T}}_{t-1}^b(k_{t-1}^1, k_{t-1}^2) \tilde{\mathcal{T}}_{t-1}^b(k_{t-1}^3, k_{t-1}^4).
\end{aligned}$$

Similarly, if  $V_{k_t^1} = \{k_t^1, k_t^4\}$  and  $V_{k_t^2} = \{k_t^2, k_t^3\}$ , we obtain

$$\begin{aligned}
& \mathbb{E} \left[ \tilde{\mathcal{T}}_t^b(k_t^1, k_t^2) \tilde{\mathcal{T}}_t^b(k_t^3, k_t^4) \middle| \mathcal{G}_{t-1}^N \right] \\
&\leq \frac{G_\infty^2}{M\Omega_{t-1}^2} \sum_{k_{t-1}^{1:2} \in [N]^2} \tilde{\mathcal{T}}_{t-1}^b(k_{t-1}^1, k_{t-1}^2) \tilde{\mathcal{T}}_{t-1}^b(k_{t-1}^2, k_{t-1}^1) + \frac{(M-1)G_\infty^2}{M\Omega_{t-1}^2} \\
&\times \sum_{k_{t-1}^{1:4} \in [N]^4} \text{M}_t[\beta_t^N(., \xi_{t-1}^{k_{t-1}^4})](\xi_{t-1}^{k_{t-1}^1}) \text{M}_t[\beta_t^N(., \xi_{t-1}^{k_{t-1}^3})](\xi_{t-1}^{k_{t-1}^2}) \tilde{\mathcal{T}}_{t-1}^b(k_{t-1}^1, k_{t-1}^2) \tilde{\mathcal{T}}_{t-1}^b(k_{t-1}^3, k_{t-1}^4).
\end{aligned}$$

Consequently,

$$\begin{aligned}
& \mathbb{E} \left[ \sum_{k_t^{1:4} \in \mathcal{I}_0^2 \cap \mathcal{S}_2^2} \tilde{\mathcal{T}}_t^b(k_t^1, k_t^2) \tilde{\mathcal{T}}_t^b(k_t^3, k_t^4) \middle| \mathcal{G}_{t-1}^N \right] \\
& \leq \frac{G_\infty^2 N(N-1)}{M\Omega_{t-1}^2} \sum_{k_{t-1}^{1:2} \in [N]^2} \left\{ \tilde{\mathcal{T}}_{t-1}^b(k_{t-1}^1, k_{t-1}^2)^2 + \tilde{\mathcal{T}}_{t-1}^b(k_{t-1}^1, k_{t-1}^2) \tilde{\mathcal{T}}_{t-1}^b(k_{t-1}^2, k_{t-1}^1) \right\} \\
& + \frac{G_\infty^2 N(N-1)(M-1)}{M\Omega_{t-1}^2} \int \sum_{k_{t-1}^{1:4} \in [N]^4} \left\{ m_t(\xi_{t-1}^{k_{t-1}^1}, x) \beta_t^N(x, \xi_{t-1}^{k_{t-1}^3}) m_t(\xi_{t-1}^{k_{t-1}^2}, y) \beta_t^N(y, \xi_{t-1}^{k_{t-1}^4}) + \right. \\
& \left. m_t(\xi_{t-1}^{k_{t-1}^1}, x) \beta_t^N(x, \xi_{t-1}^{k_{t-1}^4}) m_t(\xi_{t-1}^{k_{t-1}^2}, y) \beta_t^N(y, \xi_{t-1}^{k_{t-1}^3}) \right\} \tilde{\mathcal{T}}_{t-1}^b(k_{t-1}^1, k_{t-1}^2) \tilde{\mathcal{T}}_{t-1}^b(k_{t-1}^3, k_{t-1}^4) \nu^{\otimes 2}(dx, dy) .
\end{aligned}$$

Then, using (A1.38),

$$\begin{aligned}
& \mathbb{E} \left[ \mathcal{C}_{N,b,t}^2 \sum_{k_t^{1:4} \in \mathcal{I}_0^2 \cap \mathcal{S}_2^2} \tilde{\mathcal{T}}_t^b(k_t^1, k_t^2) \tilde{\mathcal{T}}_t^b(k_t^3, k_t^4) \middle| \mathcal{G}_{t-1}^N \right] \\
& \leq \frac{G_\infty^2 \Omega_{t-1}^2}{MN(N-1)} \mathcal{C}_{N,b,t-1}^2 \sum_{k_{t-1}^{1:2} \in [N]^2} \left\{ \tilde{\mathcal{T}}_{t-1}^b(k_{t-1}^1, k_{t-1}^2)^2 + \tilde{\mathcal{T}}_{t-1}^b(k_{t-1}^1, k_{t-1}^2) \tilde{\mathcal{T}}_{t-1}^b(k_{t-1}^2, k_{t-1}^1) \right\} \\
& + \frac{G_\infty^2 (M-1) \Omega_{t-1}^2}{MN(N-1)} \int \left\{ \tilde{\mathcal{Q}}_{b,t-1}^{N,M}(m_t(\cdot, x) \otimes m_t(\cdot, y)) \tilde{\mathcal{Q}}_{b,t-1}^{N,M}(\beta_t^N(x, \cdot) \otimes \beta_t^N(y, \cdot)) \right. \\
& \quad \left. + \tilde{\mathcal{Q}}_{b,t-1}^{N,M}(m_t(\cdot, x) \otimes m_t(\cdot, y)) \tilde{\mathcal{Q}}_{b,t-1}^{N,M}(\beta_t^N(y, \cdot) \otimes \beta_t^N(x, \cdot)) \right\} \nu^{\otimes 2}(dx, dy) ,
\end{aligned}$$

which yields (A1.34).

– If  $(k_t^1, \dots, k_t^4) \in \mathcal{I}_0^2 \cap \mathcal{S}_2^2$ . Then either  $\mathbf{k}_3 = \{k_t^1, k_t^2, k_t^3\}$  or  $\mathbf{k}_3 = \{k_t^1, k_t^2, k_t^4\}$ . Assume that  $\mathbf{k}_3 = \{k_t^1, k_t^2, k_t^3\}$ . Then  $V_{k_t^1} = \{k_t^1, k_t^4\}$ ,  $J_{k_t^1, t-1}^i = J_{k_t^4, t-1}^i$  for any  $i \in [M]$  and

$$\begin{aligned}
& \mathbb{E} \left[ \tilde{\mathcal{T}}_t^b(k_t^1, k_t^2) \tilde{\mathcal{T}}_t^b(k_t^3, k_t^4) \middle| \mathcal{G}_{t-1}^N \right] \\
& = \mathbb{E} \left[ M^{-2} \sum_{i,j \in [M]^2} \tilde{\mathcal{T}}_{t-1}^b(J_{k_t^1, t-1}^i, J_{k_t^2, t-1}^i) \tilde{\mathcal{T}}_{t-1}^b(J_{k_t^3, t-1}^j, J_{k_t^4, t-1}^j) \middle| \mathcal{G}_{t-1}^N \right] \\
& = \mathbb{E} \left[ M^{-2} \sum_{i,j \in [M]^2} \mathbb{1}_{i=j} \sum_{k_{t-1}^{1:3} \in [N]^3} \prod_{i=1}^3 \beta_t^{\text{BS}}(k_t^i, k_{t-1}^i) \tilde{\mathcal{T}}_{t-1}^b(k_{t-1}^1, k_{t-1}^2) \tilde{\mathcal{T}}_{t-1}^b(k_{t-1}^3, k_{t-1}^1) \right. \\
& \quad \left. + M^{-2} \sum_{i,j \in [M]^2} \mathbb{1}_{i \neq j} \sum_{k_{t-1}^{1:4} \in [N]^4} \prod_{\ell=1}^4 \beta_t^{\text{BS}}(k_t^\ell, k_{t-1}^\ell) \tilde{\mathcal{T}}_{t-1}^b(k_{t-1}^1, k_{t-1}^2) \tilde{\mathcal{T}}_{t-1}^b(k_{t-1}^3, k_{t-1}^4) \middle| \mathcal{G}_{t-1}^N \right] \\
& \leq \frac{G_\infty^3}{M\Omega_{t-1}^3} \sum_{k_{t-1}^{1:3} \in [N]^3} \tilde{\mathcal{T}}_{t-1}^b(k_{t-1}^1, k_{t-1}^2) \tilde{\mathcal{T}}_{t-1}^b(k_{t-1}^3, k_{t-1}^1) \\
& \quad + \frac{(M-1)G_\infty^3}{M\Omega_{t-1}^3} \sum_{k_{t-1}^{1:4} \in [N]^4} M_t[\beta_t^N(\cdot, \xi_{t-1}^{k_{t-1}^4})](\xi_{t-1}^{k_{t-1}^1}) \tilde{\mathcal{T}}_t^b(k_{t-1}^1, k_{t-1}^2) \tilde{\mathcal{T}}_t^b(k_{t-1}^3, k_{t-1}^4) .
\end{aligned}$$



The remaining combinations are treated in the exact same way and (A1.35) is obtained by using again (A1.38).

Consider now the case  $b_t = 1$  and let  $(k_t^1, \dots, k_t^4) \in \mathcal{I}_1^2 \cap \mathcal{S}_2^2$ . Then  $k_t^1 = k_t^2, k_t^3 = k_t^4$  and  $k_t^1 \neq k_t^3$ . Thus,

$$\begin{aligned}
& \mathbb{E} \left[ h(\xi_t^{k_t^1}, \xi_t^{k_t^2}) h(\xi_t^{k_t^3}, \xi_t^{k_t^4}) M^{-2} \sum_{i,j \in [M]^2} \sum_{k_{t-1}^{2,4} \in [N]^2} \mathcal{W}_{t-1}^{k_{t-1}^2} \mathcal{W}_{t-1}^{k_{t-1}^4} \right. \\
& \quad \left. \times \tilde{\mathcal{T}}_t^b(J_{k_t^1, t-1}^i, k_{t-1}^2) \tilde{\mathcal{T}}_t^b(J_{k_t^3, t-1}^j, k_{t-1}^4) \middle| \mathcal{G}_{t-1}^N \right] \\
&= \mathbb{E} \left[ h(\xi_t^{k_t^1}, \xi_t^{k_t^2}) h(\xi_t^{k_t^3}, \xi_t^{k_t^4}) \sum_{k_{t-1}^{1:4} \in [N]^4} \prod_{\ell=1}^2 \beta^{\text{BS}}(k_t^{2\ell-1}, k_{t-1}^{2\ell-1}) \mathcal{W}_{t-1}^{k_{t-1}^{2\ell-1}} \right. \\
& \quad \left. \times \tilde{\mathcal{T}}_{t-1}^b(k_{t-1}^1, k_{t-1}^2) \tilde{\mathcal{T}}_{t-1}^b(k_{t-1}^3, k_{t-1}^4) \middle| \mathcal{G}_{t-1}^N \right] \\
&= \frac{1}{\Omega_{t-1}^4} \sum_{k_{t-1}^{1:4} \in [N]^4} \tilde{\mathcal{T}}_{t-1}^b(k_{t-1}^1, k_{t-1}^2) \tilde{\mathcal{T}}_{t-1}^b(k_{t-1}^3, k_{t-1}^4) (g_{t-1}^{\otimes 2} \mathcal{M}_t^1[h])^{\otimes 2} (\xi_{t-1}^{k_{t-1}^1}, \xi_{t-1}^{k_{t-1}^2}, \xi_{t-1}^{k_{t-1}^3}, \xi_{t-1}^{k_{t-1}^4}) .
\end{aligned}$$

Since  $b_t = 1$  implies that

$$\mathcal{C}_{N,b,t}^2 = \mathcal{C}_{N,b,t-1}^2 \frac{\Omega_{t-1}^4}{N^2} ,$$

and using  $\text{Card}(\mathcal{I}_1^2 \cap \mathcal{S}_2^2) = N(N-1)$ , we get

$$\begin{aligned}
& \mathbb{E} \left[ \mathcal{C}_{N,b,t}^2 \sum_{k_t^{1:4} \in \mathcal{I}_1^2 \cap \mathcal{S}_2^2} h(\xi_t^{k_t^1}, \xi_t^{k_t^2}) h(\xi_t^{k_t^3}, \xi_t^{k_t^4}) \tilde{\mathcal{T}}_t^b(k_t^1, k_t^2) \tilde{\mathcal{T}}_t^b(k_t^3, k_t^4) \middle| \mathcal{G}_{t-1}^N \right] \\
&= \mathcal{C}_{N,b,t}^2 \sum_{k_t^{1:4} \in \mathcal{I}_1^2 \cap \mathcal{S}_2^2} \mathbb{E} \left[ h(\xi_t^{k_t^1}, \xi_t^{k_t^2}) h(\xi_t^{k_t^3}, \xi_t^{k_t^4}) M^{-2} \sum_{i,j \in [M]^2} \sum_{k_{t-1}^{2,4} \in [N]^2} \mathcal{W}_{t-1}^{k_{t-1}^2} \mathcal{W}_{t-1}^{k_{t-1}^4} \right. \\
& \quad \left. \times \tilde{\mathcal{T}}_t^b(J_{k_t^1, t-1}^i, k_{t-1}^2) \tilde{\mathcal{T}}_t^b(J_{k_t^3, t-1}^j, k_{t-1}^4) \middle| \mathcal{G}_{t-1}^N \right] \\
&= \frac{N-1}{N} \tilde{\mathcal{Q}}_{b,t-1}^{N,M} (g_{t-1}^{\otimes 2} \mathcal{M}_t^1[h])^2 ,
\end{aligned}$$

which proves (A1.36).

Let  $(k_t^1, \dots, k_t^4) \in \mathcal{I}_1^2 \cap \mathcal{S}_2^1$ . Then  $\mathbf{k}_1 = \{k_t^1\}$ ,  $k_t^1 = k_t^2 = k_t^3 = k_t^4$  and  $J_{k_t^1, t-1}^i = J_{k_t^3, t-1}^i$ . Hence,

$$\begin{aligned}
& \mathbb{E} \left[ M^{-2} \sum_{i,j \in [M]^2} \sum_{k_{t-1}^{2,4} \in [N]^2} \mathcal{W}_{t-1}^{k_{t-1}^2} \mathcal{W}_{t-1}^{k_{t-1}^4} \tilde{\mathcal{T}}_{t-1}^b(J_{k_t^1, t-1}^i, k_{t-1}^2) \tilde{\mathcal{T}}_{t-1}^b(J_{k_t^3, t-1}^j, k_{t-1}^4) | \mathcal{G}_{t-1}^N \right] \\
&= \mathbb{E} \left[ \frac{1}{M} \sum_{k_{t-1}^{1,2,4} \in [N]^3} \beta_t^{\text{BS}}(k_t^1, k_{t-1}^1) \mathcal{W}_{t-1}^{k_{t-1}^2} \mathcal{W}_{t-1}^{k_{t-1}^4} \tilde{\mathcal{T}}_{t-1}^b(k_{t-1}^1, k_{t-1}^2) \tilde{\mathcal{T}}_{t-1}^b(k_{t-1}^1, k_{t-1}^4) \right. \\
&\quad \left. + \frac{M-1}{M} \sum_{k_{t-1}^{1,4} \in [N]^2} \prod_{\ell=1}^2 \beta_t^{\text{BS}}(k_t^{2\ell-1}, k_{t-1}^{2\ell-1}) \mathcal{W}_{t-1}^{k_{t-1}^{2\ell}} \tilde{\mathcal{T}}_{t-1}^b(k_{t-1}^1, k_{t-1}^{2\ell-1}) \tilde{\mathcal{T}}_{t-1}^b(k_{t-1}^3, k_{t-1}^{2\ell-1}) | \mathcal{G}_{t-1}^N \right] \\
&= \frac{G_\infty^3}{M \Omega_{t-1}^3} \sum_{k_{t-1}^{1,2,4} \in [N]^3} \tilde{\mathcal{T}}_{t-1}^b(k_{t-1}^1, k_{t-1}^2) \tilde{\mathcal{T}}_{t-1}^b(k_{t-1}^1, k_{t-1}^4) \\
&\quad + \frac{G_\infty^3 (M-1)}{M \Omega_{t-1}^3} \int \tilde{\mathcal{Q}}_{b,t-1}^{N,M}(m_t(\cdot, x) \otimes \mathbf{1}) \tilde{\mathcal{Q}}_{b,t-1}^{N,M}(\beta_t^N(x, \cdot) \otimes \mathbf{1}) \nu(dx) .
\end{aligned}$$

This yields

$$\begin{aligned}
& \mathbb{E} \left[ \mathcal{C}_{N,b,t}^2 \sum_{k_t^{1,4} \in \mathcal{I}_1^2 \cap \mathcal{S}_2^1} \tilde{\mathcal{T}}_t^b(k_t^1, k_t^2) \tilde{\mathcal{T}}_t^b(k_t^3, k_t^4) | \mathcal{G}_{t-1}^N \right] \\
&\leq \frac{G_\infty^3 \Omega_{t-1}}{MN} \mathcal{C}_{N,b,t-1}^2 \sum_{k_{t-1}^{1,2,4} \in [N]^3} \tilde{\mathcal{T}}_{t-1}^b(k_{t-1}^1, k_{t-1}^2) \tilde{\mathcal{T}}_{t-1}^b(k_{t-1}^1, k_{t-1}^4) \\
&\quad + \frac{G_\infty^3 (M-1) \Omega_{t-1}}{MN} \int \tilde{\mathcal{Q}}_{b,t-1}^{N,M}(m_t(\cdot, x) \otimes \mathbf{1}) \tilde{\mathcal{Q}}_{b,t-1}^{N,M}(\beta_t^N(x, \cdot) \otimes \mathbf{1}) \nu(dx) ,
\end{aligned}$$

which in turn proves (A1.37).  $\square$

**Proposition A.11.** *Let (A2 : 4) hold. Let  $t \geq 0$ . For any  $b \in \mathcal{B}_t$*

(i) *If  $b_t = 0$ , then for  $p \in \{2, 3\}$ ,*

$$\lim_{N \rightarrow \infty} \mathbb{E} \left[ \mathcal{C}_{N,b,t}^2 \sum_{k_t^{1,4} \in \mathcal{I}_0^2 \cap \mathcal{S}_2^p} \tilde{\mathcal{T}}_t^b(k_t^1, k_t^2) \tilde{\mathcal{T}}_t^b(k_t^3, k_t^4) \right] = 0 . \quad (\text{A1.39})$$

(ii) *If  $b_t = 1$ , then*

$$\lim_{N \rightarrow \infty} \mathbb{E} \left[ \mathcal{C}_{N,b,t}^2 \sum_{k_t^{1,4} \in \mathcal{I}_1^2 \cap \mathcal{S}_2^1} \tilde{\mathcal{T}}_t^b(k_t^1, k_t^2) \tilde{\mathcal{T}}_t^b(k_t^3, k_t^4) \right] = 0 . \quad (\text{A1.40})$$

Additionally, if (A5) holds then the rate of convergence is  $\mathcal{O}(N^{-1})$ .

*Proof.* We prove (i)-(ii) simultaneously by induction. Let  $t = 0$  and  $b \in \mathcal{B}_0$ . By definition,  $\tilde{\mathcal{T}}_0^b = \mathcal{T}_0^b$  and, if  $b_0 = 0$ ,

$$\mathcal{C}_{N,b,0}^2 \sum_{k_0^{1:4} \in \mathcal{I}_0^2 \cap \mathcal{S}_2^2} \tilde{\mathcal{T}}_0^b(k_0^1, k_0^2) \tilde{\mathcal{T}}_0^b(k_0^3, k_0^4) = \frac{2}{N^2(N-1)^2} \sum_{i,j \in [N]^2} \mathbb{1}_{i \neq j} = \mathcal{O}(N^{-2}), \quad (\text{A1.41})$$

$$\mathcal{C}_{N,b,0}^2 \sum_{k_0^{1:4} \in \mathcal{I}_0^2 \cap \mathcal{S}_2^3} \tilde{\mathcal{T}}_0^b(k_0^1, k_0^2) \tilde{\mathcal{T}}_0^b(k_0^3, k_0^4) = \frac{4}{N^2(N-1)^2} \sum_{i,j,k \in [N]^3} \mathbb{1}_{i \neq j} \mathbb{1}_{i \neq k} = \mathcal{O}(N^{-1}). \quad (\text{A1.42})$$

If  $b_0 = 1$ , then

$$\mathcal{C}_{N,b,0}^2 \sum_{k_0^{1:4} \in \mathcal{I}_1^2 \cap \mathcal{S}_2^1} \tilde{\mathcal{T}}_0^b(k_0^1, k_0^2) \tilde{\mathcal{T}}_t^b(k_0^3, k_0^4) = \frac{1}{N^2} \sum_{i,j \in [N]^2} \mathbb{1}_{i=j} = \mathcal{O}(N^{-1}). \quad (\text{A1.43})$$

Let  $t > 0$  and assume that both (i)-(ii) hold at  $t-1$ . Define

$$\begin{aligned} D_{1,b}^N &= \mathbb{E} \left[ \mathcal{C}_{N,b,t-1}^2 \sum_{k_{t-1}^{1,2,4} \in [N]^3} \tilde{\mathcal{T}}_{t-1}^b(k_{t-1}^1, k_{t-1}^2) \tilde{\mathcal{T}}_{t-1}^b(k_{t-1}^1, k_{t-1}^4) \right], \\ D_{2,b}^N &= \mathbb{E} \left[ \mathcal{C}_{N,b,t-1}^2 \sum_{k_{t-1}^{1:2} \in [N]^2} \mathsf{T}_{t-1,b}^{(1)}(k_{t-1}^1, k_{t-1}^2) \right], \\ D_{3,b}^N &= \mathbb{E} \left[ \mathcal{C}_{N,b,t-1}^2 \sum_{k_{t-1}^{1:3} \in [N]^3} \mathsf{T}_{t-1,b}^{(2)}(k_{t-1}^1, k_{t-1}^2, k_{t-1}^3) \right], \end{aligned}$$

where  $\mathsf{T}_{t-1,b}^{(1)}$  and  $\mathsf{T}_{t-1,b}^{(2)}$  are defined in Proposition A.10. If  $b_t = 0$ , then by (i) in Proposition A.10, using that  $\Omega_{t-1} \leq NG_\infty$ ,

$$\begin{aligned} \mathbb{E} \left[ \mathcal{C}_{N,b,t}^2 \sum_{k_t^{1:4} \in \mathcal{I}_0^2 \cap \mathcal{S}_2^2} \tilde{\mathcal{T}}_t^b(k_t^1, k_t^2) \tilde{\mathcal{T}}_t^b(k_t^3, k_t^4) \right] \\ \leq \mathbb{E} \left[ \frac{\Omega_{t-1}^2 G_\infty^2 (M-1)}{MN(N-1)} \int \tilde{\Upsilon}_{b,t}^{N,M}(x, y) \nu^{\otimes 2}(\mathrm{d}x, \mathrm{d}y) \right] + 2 \frac{G_\infty^4}{M} D_{2,b}^N, \quad (\text{A1.44}) \end{aligned}$$

and

$$\begin{aligned} \mathbb{E} \left[ \mathcal{C}_{N,b,t}^2 \sum_{k_t^{1:4} \in \mathcal{I}_0^2 \cap \mathcal{S}_2^3} \tilde{\mathcal{T}}_t^b(k_t^1, k_t^2) \tilde{\mathcal{T}}_t^b(k_t^3, k_t^4) \right] \\ \leq \mathbb{E} \left[ \frac{\Omega_{t-1} G_\infty^3 (M-1)(N-2)}{MN(N-1)} \int \tilde{\Theta}_{b,t}^{N,M}(x) \nu(\mathrm{d}x) \right] + \frac{G_\infty^4}{M} D_{3,b}^N, \quad (\text{A1.45}) \end{aligned}$$

where  $\tilde{\Theta}_{b,t}^{N,M}$  and  $\tilde{\Upsilon}_{b,t}^{N,M}$  are defined in Proposition A.10. We deal first with  $D_{2,b}^N$  and  $D_{3,b}^N$ . If  $b_{t-1} = 0$ , then by definition of  $\tilde{\mathcal{T}}_{t-1}^b$  in (4.4),

$$\begin{aligned} D_{2,b}^N &= \mathbb{E} \left[ \mathcal{C}_{N,b,t-1}^2 \sum_{k_{t-1}^{1:4} \in \mathcal{I}_0^2 \cap \mathcal{S}_2^2} \tilde{\mathcal{T}}_{t-1}^b(k_{t-1}^1, k_{t-1}^2) \tilde{\mathcal{T}}_{t-1}^b(k_{t-1}^3, k_{t-1}^4) \right], \\ D_{3,b}^N &= \mathbb{E} \left[ \mathcal{C}_{N,b,t-1}^2 \sum_{k_{t-1}^{1:4} \in \mathcal{I}_0^2 \cap \mathcal{S}_2^3} \tilde{\mathcal{T}}_{t-1}^b(k_{t-1}^1, k_{t-1}^2) \tilde{\mathcal{T}}_{t-1}^b(k_{t-1}^3, k_{t-1}^4) \right]. \quad (\text{A1.46}) \end{aligned}$$

If  $b_{t-1} = 1$ ,

$$\begin{aligned} D_{2,b}^N &= 2\mathbb{E}\left[\mathcal{C}_{N,b,t-1}^2 \sum_{k_{t-1}^{1:4} \in \mathcal{I}_1^2 \cap \mathcal{S}_2^1} \tilde{\mathcal{T}}_{t-1}^b(k_{t-1}^1, k_{t-1}^2) \tilde{\mathcal{T}}_{t-1}^b(k_{t-1}^3, k_{t-1}^4)\right], \\ D_{3,b}^N &= 4\mathbb{E}\left[\mathcal{C}_{N,b,t-1}^2 \sum_{k_{t-1}^{1:4} \in \mathcal{I}_1^2 \cap \mathcal{S}_2^1} \tilde{\mathcal{T}}_{t-1}^b(k_{t-1}^1, k_{t-1}^2) \tilde{\mathcal{T}}_{t-1}^b(k_{t-1}^3, k_{t-1}^4)\right]. \end{aligned} \quad (\text{A1.47})$$

In all cases, by the induction hypothesis we get for any  $b \in \mathcal{B}_t$  with  $b_t = 0$ ,

$$\lim_{N \rightarrow \infty} D_{2,b}^N = 0, \quad \lim_{N \rightarrow \infty} D_{3,b}^N = 0.$$

Regarding the first terms in the r.h.s. of inequalities (A1.44)-(A1.45), they go to zero when  $N$  goes to infinity since they are, up to the constant  $(M-1)/M \leq 1$ , the *PaRIS* counterpart of  $B_N$  (A1.12) in the proof of Theorem 4.4 and are treated in the exact same way since  $\sup_{N \in \mathbb{N}} \mathbb{E}[\tilde{\mathcal{Q}}_{b,t}^{N,M}(\mathbf{1})^3] < \infty$  by Proposition A.12. If  $b_t = 1$ , then, by Proposition A.10, using that  $\Omega_{t-1} \leq NG_\infty$ ,

$$\begin{aligned} &\mathbb{E}\left[\mathcal{C}_{N,b,t}^2 \sum_{k_t^{1:4} \in \mathcal{I}_1^2 \cap \mathcal{S}_2^1} \tilde{\mathcal{T}}_t^b(k_t^1, k_t^2) \tilde{\mathcal{T}}_t^b(k_t^3, k_t^4)\right] \\ &\leq \mathbb{E}\left[\frac{\Omega_{t-1} G_\infty^3}{N} \int \tilde{\mathcal{Q}}_{b,t-1}^{N,M}(m_t(\cdot, x) \otimes \mathbf{1}) \tilde{\mathcal{Q}}_{b,t-1}^{N,M}(\beta_t^N(x, \cdot) \otimes \mathbf{1}) \nu(dx)\right] + \frac{G_\infty^4}{M} D_{1,b}^N. \end{aligned}$$

The first term goes to zero when  $N$  goes to infinity similarly to the case  $b_t = 0$ . As for the second term, if  $b_{t-1} = 0$  then by definition of  $D_{1,b}^N$ ,

$$0 \leq D_{1,b}^N \leq D_{3,b}^N,$$

and if  $b_{t-1} = 1$ ,

$$0 \leq D_{1,b}^N = \mathbb{E}\left[\mathcal{C}_{N,b,t}^2 \sum_{k_{t-1}^{1:2} \in [N]^2} \tilde{\mathcal{T}}_{t-1}^b(k_{t-1}^1, k_{t-1}^2)^2\right] \leq D_{2,b}^N.$$

Hence, in both cases  $D_{1,b}^N$  goes to zero by (A.10). This ends the proof of the first claim.

Now assume that (A5) holds also. We proceed by induction. At  $t = 0$  the rate of convergence is  $\mathcal{O}(N^{-1})$  by (A1.41)-(A1.42) and (A1.43). Let  $t > 0$  and assume that the rate of convergence in (i) and (ii) at  $t-1$  is  $\mathcal{O}(N^{-1})$ . Assume that  $b_t = 0$ . By the strong mixing assumption we have that

$$\beta_t^N(x, y) \leq \frac{G_\infty \sigma_+}{\sigma_- \Omega_{t-1}}, \quad \forall (x, y) \in \mathbb{X}^2. \quad (\text{A1.48})$$

Using for example that

$$\int \tilde{\mathcal{Q}}_{b,t}^{N,M}(m_t(\cdot, x) \otimes \mathbf{1}) \nu(dx) = \int \tilde{\mathcal{Q}}_{b,t}^{N,M}(m_t(\cdot, x) \otimes m_t(\cdot, y)) \nu^{\otimes 2}(dx, dy) = \tilde{\mathcal{Q}}_{b,t}^{N,M}(\mathbf{1} \otimes \mathbf{1}),$$

and then bounding  $\beta_t^N$  using (A1.48) we get that

$$\begin{aligned} &\mathbb{E}\left[\frac{\Omega_{t-1}^2 G_\infty^2 (M-1)}{MN(N-1)} \int \tilde{\Upsilon}_{b,t}^{N,M}(x, y) \nu^{\otimes 2}(dx, dy)\right] \\ &\leq \frac{G_\infty^4 \sigma_+^2 (M-1)}{\sigma_-^2 MN(N-1)} \|\tilde{\mathcal{Q}}_{b,t-1}^{N,M}(\mathbf{1} \otimes \mathbf{1})\|_2^2, \end{aligned}$$

and

$$\mathbb{E} \left[ \frac{\Omega_{t-1} G_\infty^3 (M-1)(N-2)}{MN(N-1)} \int \tilde{\Theta}_{b,t}^{N,M}(x) \nu(dx) \right] \leq \frac{4\sigma_+ G_\infty^4 (M-1)(N-2)}{\sigma_- MN(N-1)} \|\tilde{\mathcal{Q}}_{b,t-1}^{N,M}(\mathbf{1} \otimes \mathbf{1})\|_2^2,$$

where both bounds are  $\mathcal{O}(N^{-1})$  by Proposition A.12. Going back to (A1.44)-(A1.45), we obtain

$$\mathbb{E} \left[ \mathcal{C}_{N,b,t}^2 \sum_{k_t^{1:4} \in \mathcal{I}_0^2 \cap \mathcal{S}_2^2} \tilde{\mathcal{T}}_t^b(k_t^1, k_t^2) \tilde{\mathcal{T}}_t^b(k_t^3, k_t^4) \right] \leq \frac{2G_\infty^4 \sigma_+^2 (M-1)}{\sigma_-^2 MN(N-1)} \|\tilde{\mathcal{Q}}_{b,t-1}^{N,M}(\mathbf{1} \otimes \mathbf{1})\|_2^2 + 2 \frac{G_\infty^4}{M} D_{2,b}^N,$$

and

$$\begin{aligned} \mathbb{E} \left[ \mathcal{C}_{N,b,t}^2 \sum_{k_t^{1:4} \in \mathcal{I}_0^2 \cap \mathcal{S}_2^3} \tilde{\mathcal{T}}_t^b(k_t^1, k_t^2) \tilde{\mathcal{T}}_t^b(k_t^3, k_t^4) \right] \\ \leq \frac{4\sigma_+ G_\infty^4 (M-1)(N-2)}{\sigma_- MN(N-1)} \|\tilde{\mathcal{Q}}_{b,t-1}^{N,M}(\mathbf{1} \otimes \mathbf{1})\|_2^2 + \frac{G_\infty^4}{M} D_{3,b}^N. \end{aligned}$$

By (A1.46)-(A1.47) and the induction hypothesis, we get that  $D_{2,b}^N$  and  $D_{3,b}^N$  are both  $\mathcal{O}(N^{-1})$ . Finally, applying Proposition A.12 we get  $\mathcal{O}(N^{-1})$  upper bounds. This ends the proof for the case  $b_t = 0$ . The case  $b_t = 1$  follows the same steps.  $\square$

**Proposition A.12.** Assume that (A2) holds. For all  $M > 1$ ,  $t \in \mathbb{N}$  and  $b \in \mathcal{B}_t$ ,

$$\sup_{N \in \mathbb{N}} \|\tilde{\mathcal{Q}}_{b,t}^{N,M}(\mathbf{1})\|_3 < \infty. \quad (\text{A1.49})$$

*Proof.* We proceed by induction on  $t \in \mathbb{N}$ . Assume for now that  $N \geq 6$  and  $M \geq 2$ .

Since  $\tilde{\mathcal{Q}}_{b,0}^{N,M}(h) = \mathcal{Q}_{b,0}^{N,\text{BS}}(h)$  for any  $h$ , the case  $t = 0$  follows from Proposition A.4. Let  $t > 0$  and assume that (A1.49) holds at  $t-1$ . We only treat the case  $b_t = 0$ . The proof for the case  $b_t = 1$  follows using the same steps. Since we have that

$$\begin{aligned} \|\tilde{\mathcal{Q}}_{b,t}^{N,M}(\mathbf{1})\|_3^3 &= \mathbb{E} \left[ \mathcal{C}_{N,b,t}^3 \sum_{k_t^{1:6} \in [N]^6} \prod_{\ell=1}^3 \tilde{\mathcal{T}}_t^b(k_t^{2\ell-1}, k_t^{2\ell}) \right] \\ &= \mathbb{E} \left[ \frac{\mathcal{C}_{N,b,t}^3}{M^3} \sum_{k_t^{1:6} \in \mathcal{I}_0^3} \sum_{i^{1:3} \in [M]^3} \prod_{\ell=1}^3 \tilde{\mathcal{T}}_{t-1}^b(J_{k_t^{2\ell-1}, t-1}^{i^\ell}, J_{k_t^{2\ell}, t-1}^{i^\ell}) \right], \end{aligned} \quad (\text{A1.50})$$

the proof proceeds by (i) splitting the sum in three parts with respect to the cardinal of the triplet  $(i^1, i^2, i^3) \in [M]^3$ , and (ii) bounding each term by  $\|\tilde{\mathcal{Q}}_{b,t-1}^{N,M}(\mathbf{1})\|_3^3$  up to a constant independent of  $N$ .

Let  $(i^1, i^2, i^3) \in [M]^3$ . If  $\text{Card}(\{i^1, i^2, i^3\}) = 3$ , then for all  $k_t^{1:6} \in [N]^6$ ,  $(J_{k_t^{2\ell-1}, t-1}^{i^\ell}, J_{k_t^{2\ell}, t-1}^{i^\ell})_{\ell \in [1:3]}$

are mutually independent conditionally on  $\mathcal{G}_{t-1}^N \vee \xi_t^{1:N}$ , defined in (A1.30), and

$$\begin{aligned}
& \mathbb{E} \left[ \prod_{\ell=1}^3 \tilde{\mathcal{T}}_{t-1}^b (J_{k_t^{2\ell-1}, t-1}^{i^\ell}, J_{k_t^{2\ell}, t-1}^{i^\ell}) \middle| \mathcal{G}_{t-1}^N \right] \\
&= \mathbb{E} \left[ \prod_{\ell=1}^3 \mathbb{E} \left[ \tilde{\mathcal{T}}_{t-1}^b (J_{k_t^{2\ell-1}, t-1}^{i^\ell}, J_{k_t^{2\ell}, t-1}^{i^\ell}) \middle| \mathcal{G}_{t-1}^N \vee \xi_t^{1:N} \right] \middle| \mathcal{G}_{t-1}^N \right] \\
&= \mathbb{E} \left[ \prod_{\ell=1}^3 \sum_{k_{t-1}^{2\ell-1:2\ell} \in [N]^2} \beta_t^{\text{BS}}(k_t^{2\ell-1}, k_{t-1}^{2\ell-1}) \beta_t^{\text{BS}}(k_t^{2\ell}, k_{t-1}^{2\ell}) \tilde{\mathcal{T}}_{t-1}^b(k_{t-1}^{2\ell-1}, k_{t-1}^{2\ell}) \middle| \mathcal{G}_{t-1}^N \right] \\
&= \sum_{k_{t-1}^{1:6} \in [N]^6} \prod_{\ell=1}^3 \tilde{\mathcal{T}}_{t-1}^b(k_{t-1}^{2\ell-1}, k_{t-1}^{2\ell}) \mathbb{E} \left[ \prod_{n=1}^6 \beta_t^{\text{BS}}(k_t^n, k_{t-1}^n) \middle| \mathcal{F}_{t-1}^N \right].
\end{aligned}$$

Hence, by Proposition A.6, Lemma A.5 and similarly to the proof of Proposition A.4,

$$\begin{aligned}
& \sum_{k_t^{1:6} \in \mathcal{I}_0^3} \mathbb{E} \left[ \prod_{\ell=1}^3 \tilde{\mathcal{T}}_{t-1}^b (J_{k_t^{2\ell-1}, t-1}^{i^\ell}, J_{k_t^{2\ell}, t-1}^{i^\ell}) \middle| \mathcal{G}_{t-1}^N \right] \\
&= \sum_{k_{t-1}^{1:6} \in [N]^6} \prod_{\ell=1}^3 \tilde{\mathcal{T}}_{t-1}^b(k_{t-1}^{2\ell-1}, k_{t-1}^{2\ell}) \sum_{k_t^{1:6} \in \mathcal{I}_0^3} \mathbb{E} \left[ \prod_{n=1}^6 \beta_t^{\text{BS}}(k_t^n, k_{t-1}^n) \middle| \mathcal{F}_{t-1}^N \right] \\
&\leq \sum_{k_{t-1}^{1:6} \in [N]^6} \prod_{\ell=1}^3 \tilde{\mathcal{T}}_{t-1}^b(k_{t-1}^{2\ell-1}, k_{t-1}^{2\ell}) \sum_{p=2}^6 \sum_{k_t^{1:6} \in \mathcal{I}_0^3 \cap \mathcal{S}_3^p} \frac{G_\infty^p}{\Omega_{t-1}^p} \\
&\lesssim \left( \sum_{p=2}^6 \frac{N^p}{\Omega_{t-1}^p} \right) \sum_{k_{t-1}^{1:6} \in [N]^6} \prod_{\ell=1}^3 \tilde{\mathcal{T}}_{t-1}^b(k_{t-1}^{2\ell-1}, k_{t-1}^{2\ell}),
\end{aligned}$$

where  $\lesssim$  means less than or equal up to a multiplicative constant independent of  $N$ . Consequently,

$$\begin{aligned}
& \mathbb{E} \left[ \frac{\mathcal{C}_{N,b,t}^3}{M^3} \sum_{k_t^{1:6} \in \mathcal{I}_0^3} \sum_{i^{1:3} \in [M]^3} \mathbb{1}_{i^1 \neq i^2 \neq i^3} \prod_{\ell=1}^3 \tilde{\mathcal{T}}_{t-1}^b (J_{k_t^{2\ell-1}, t-1}^{i^\ell}, J_{k_t^{2\ell}, t-1}^{i^\ell}) \middle| \mathcal{G}_{t-1}^N \right] \\
&= \frac{\mathcal{C}_{N,b,t}^3}{M^3} \sum_{k_t^{1:6} \in \mathcal{I}_0^3} \sum_{i^{1:3} \in [M]^3} \mathbb{1}_{i^1 \neq i^2 \neq i^3} \mathbb{E} \left[ \prod_{\ell=1}^3 \tilde{\mathcal{T}}_{t-1}^b (J_{k_t^{2\ell-1}, t-1}^{i^\ell}, J_{k_t^{2\ell}, t-1}^{i^\ell}) \middle| \mathcal{G}_{t-1}^N \right] \\
&\lesssim \frac{M(M-1)(M-2)}{M^3} \left( \sum_{p=2}^6 \frac{N^p}{\Omega_{t-1}^p} \right) \mathcal{C}_{N,b,t}^3 \sum_{k_{t-1}^{1:6} \in [N]^6} \prod_{\ell=1}^3 \tilde{\mathcal{T}}_{t-1}^b(k_{t-1}^{2\ell-1}, k_{t-1}^{2\ell}),
\end{aligned}$$

and

$$\begin{aligned}
\mathcal{C}_{N,b,t}^3 \left( \sum_{p=2}^6 \frac{N^p}{\Omega_{t-1}^p} \right) &= \mathcal{C}_{N,b,t-1}^3 \left( \sum_{p=2}^6 \frac{N^p \Omega_{t-1}^6}{\Omega_{t-1}^p N^3 (N-1)^3} \right) \\
&\leq \mathcal{C}_{N,b,t-1}^3 \left( \sum_{p=2}^6 \frac{G_\infty^{6-p} N^6}{N^3 (N-1)^3} \right) \lesssim \mathcal{C}_{N,b,t-1}^3.
\end{aligned}$$

Therefore,

$$\mathbb{E} \left[ \frac{\mathcal{C}_{N,b,t}^3}{M^3} \sum_{k_t^{1:6} \in [N]^6} \sum_{i^{1:3} \in [M]^3} \mathbb{1}_{i^1 \neq i^2 \neq i^3} \prod_{\ell=1}^3 \tilde{\mathcal{T}}_{t-1}^b(J_{k_t^{2\ell-1}, t-1}^{i^\ell}, J_{k_t^{2\ell}, t-1}^{i^\ell}) \right] \lesssim \|\tilde{\mathcal{Q}}_{b,t-1}^{N,M}(\mathbf{1})\|_3^3. \quad (\text{A1.51})$$

Assume now that  $\text{Card}(\{i^1, i^2, i^3\}) = 2$  and that  $i^1 = i^2$ . For all  $(k_t^1, \dots, k_t^6) \in [N]^6$ , conditionally on  $\mathcal{G}_{t-1}^N \vee \xi_t^{1:N}$ ,  $\tilde{\mathcal{T}}_{t-1}^b(J_{k_t^{i^3}, t-1}^{i^3}, J_{k_t^{i^6}, t-1}^{i^6})$  is independent from  $\tilde{\mathcal{T}}_{t-1}^b(J_{k_t^{i^1}, t-1}^{i^1}, J_{k_t^{i^2}, t-1}^{i^2})$  and  $\tilde{\mathcal{T}}_{t-1}^b(J_{k_t^{i^3}, t-1}^{i^3}, J_{k_t^{i^4}, t-1}^{i^4})$ , hence, using (A1.5),

$$\begin{aligned} & \sum_{k_t^{1:6} \in \mathcal{I}_0^3} \mathbb{E} \left[ \prod_{\ell=1}^3 \tilde{\mathcal{T}}_{t-1}^b(J_{k_t^{2\ell-1}, t-1}^{i^\ell}, J_{k_t^{2\ell}, t-1}^{i^\ell}) \middle| \mathcal{G}_{t-1}^N \vee \xi_t^{1:N} \right] \\ &= \left\{ \sum_{k_t^{1:4} \in \mathcal{I}_0^2} \mathbb{E} \left[ \prod_{\ell=1}^2 \tilde{\mathcal{T}}_{t-1}^b(J_{k_t^{2\ell-1}, t-1}^{i^\ell}, J_{k_t^{2\ell}, t-1}^{i^\ell}) \middle| \mathcal{G}_{t-1}^N \vee \xi_t^{1:N} \right] \right\} \\ & \quad \times \left\{ \sum_{k_t^{5:6} \in \mathcal{I}_0^1} \mathbb{E} \left[ \tilde{\mathcal{T}}_{t-1}^b(J_{k_t^{i^3}, t-1}^{i^3}, J_{k_t^{i^6}, t-1}^{i^6}) \middle| \mathcal{G}_{t-1}^N \vee \xi_t^{1:N} \right] \right\} \\ &= (F_{2,b}^N + F_{3,b}^N + F_{4,b}^N) F_{1,b}^N, \end{aligned} \quad (\text{A1.52})$$

with

$$\begin{aligned} F_{1,b}^N &:= \sum_{k_t^{5:6} \in \mathcal{I}_0^1} \mathbb{E} \left[ \tilde{\mathcal{T}}_{t-1}^b(J_{k_t^{i^3}, t-1}^{i^3}, J_{k_t^{i^6}, t-1}^{i^6}) \middle| \mathcal{G}_{t-1}^N \vee \xi_t^{1:N} \right] \\ &= \sum_{k_t^{5:6} \in [N]^2} \mathbb{1}_{k_t^5 \neq k_t^6} \sum_{k_{t-1}^{5:6} \in [N]^2} \beta_t^{\text{BS}}(k_t^5, k_{t-1}^5) \beta_t^{\text{BS}}(k_t^6, k_{t-1}^6) \tilde{\mathcal{T}}_{t-1}^b(k_{t-1}^5, k_{t-1}^6), \end{aligned}$$

and

$$\begin{aligned} & \sum_{k_t^{1:4} \in \mathcal{I}_0^2} \mathbb{E} \left[ \prod_{\ell=1}^2 \tilde{\mathcal{T}}_{t-1}^b(J_{k_t^{2\ell-1}, t-1}^{i^\ell}, J_{k_t^{2\ell}, t-1}^{i^\ell}) \middle| \mathcal{G}_{t-1}^N \vee \xi_t^{1:N} \right] \\ &= \sum_{p=2}^4 \sum_{k_t^{1:4} \in \mathcal{I}_0^2 \cap \mathcal{S}_2^p} \mathbb{E} \left[ \prod_{\ell=1}^2 \tilde{\mathcal{T}}_{t-1}^b(J_{k_t^{2\ell-1}, t-1}^{i^\ell}, J_{k_t^{2\ell}, t-1}^{i^\ell}) \middle| \mathcal{G}_{t-1}^N \vee \xi_t^{1:N} \right] \\ &= F_{2,b}^N + F_{3,b}^N + F_{4,b}^N, \end{aligned}$$

where

$$\begin{aligned} F_{2,b}^N &:= \sum_{k_t^{1:2} \in [N]^2} \mathbb{1}_{k_t^2 \neq k_t^1} \sum_{k_{t-1}^{1:2} \in [N]^2} \beta_t^{\text{BS}}(k_t^1, k_{t-1}^1) \beta_t^{\text{BS}}(k_t^2, k_{t-1}^2) \Upsilon_{t-1,b}^{(1)}(k_{t-1}^1, k_{t-1}^2) \\ F_{3,b}^N &:= \sum_{k_t^{1:3} \in [N]^3} \mathbb{1}_{k_t^1 \neq k_t^2 \neq k_t^3} \sum_{k_{t-1}^{1:3} \in [N]^3} \prod_{\ell=1}^3 \beta_t^{\text{BS}}(k_t^\ell, k_{t-1}^\ell) \Upsilon_{t-1,b}^{(2)}(k_{t-1}^1, k_{t-1}^2, k_{t-1}^3) \\ F_{4,b}^N &:= \sum_{k_t^{1:4} \in [N]^4} \mathbb{1}_{k_t^1 \neq k_t^2 \neq k_t^3 \neq k_t^4} \sum_{k_{t-1}^{1:4} \in [N]^4} \prod_{\ell=1}^4 \beta_t^{\text{BS}}(k_t^\ell, k_{t-1}^\ell) \tilde{\mathcal{T}}_{t-1}^b(k_{t-1}^1, k_{t-1}^2) \tilde{\mathcal{T}}_{t-1}^b(k_{t-1}^3, k_{t-1}^4), \end{aligned}$$

where  $\mathsf{T}_{t-1,b}^{(1)}$  and  $\mathsf{T}_{t-1,b}^{(2)}$  are defined in Proposition A.10. We now upperbound each  $\mathbb{E}[F_{1,b}^N F_{i,b}^N | \mathcal{G}_{t-1}^N]$  for  $i \in [2 : 4]$ .

Consider first the case  $\mathbb{E}[F_{1,b}^N F_{4,b}^N | \mathcal{G}_{t-1}^N]$ . Let  $S_6 := \{k_t^{1:6} \in [N]^6 : k_t^1 \neq k_t^2 \neq k_t^3 \neq k_t^4, k_t^5 \neq k_t^6\}$ . Then,  $S_6 \subset (\mathcal{I}_0^3 \cap \mathcal{S}_3^4) \sqcup (\mathcal{I}_0^3 \cap \mathcal{S}_3^5) \sqcup (\mathcal{I}_0^3 \cap \mathcal{S}_3^6)$  and

$$\begin{aligned} \mathbb{E}[F_{1,b}^N F_{4,b}^N | \mathcal{G}_{t-1}^N] &= \sum_{k_{t-1}^{1:6} \in [N]^6} \prod_{\ell=1}^3 \tilde{\mathcal{T}}_{t-1}^b(k_{t-1}^{2\ell-1}, k_{t-1}^{2\ell}) \sum_{k_t^{1:6} \in S_6} \mathbb{E} \left[ \prod_{n=1}^6 \beta_t^{\text{BS}}(k_t^n, k_{t-1}^n) \middle| \mathcal{G}_{t-1}^N \right] \\ &\leq \sum_{k_{t-1}^{1:6} \in [N]^6} \prod_{\ell=1}^3 \tilde{\mathcal{T}}_{t-1}^b(k_{t-1}^{2\ell-1}, k_{t-1}^{2\ell}) \sum_{p=4}^6 \sum_{k_t^{1:6} \in \mathcal{I}_0^3 \cap \mathcal{S}_3^p} \mathbb{E} \left[ \prod_{n=1}^6 \beta_t^{\text{BS}}(k_t^n, k_{t-1}^n) \middle| \mathcal{G}_{t-1}^N \right] \\ &\lesssim \left( \sum_{p=4}^6 \frac{N^p G_\infty^p}{\Omega_{t-1}^p} \right) \sum_{k_{t-1}^{1:6} \in [N]^6} \prod_{\ell=1}^3 \tilde{\mathcal{T}}_{t-1}^b(k_{t-1}^{2\ell-1}, k_{t-1}^{2\ell}), \end{aligned} \quad (\text{A1.53})$$

by Proposition A.6 and Lemma A.5.

Consider now the case  $\mathbb{E}[F_{1,b}^N F_{3,b}^N | \mathcal{G}_{t-1}^N]$  and define  $S_5 := \{k_t^{1:5} \in [N]^5 : k_t^1 \neq k_t^2 \neq k_t^3, k_t^4 \neq k_t^5\}$ . Then,

$$\begin{aligned} &\mathbb{E}[F_{1,b}^N F_{3,b}^N | \mathcal{G}_{t-1}^N] \\ &= \sum_{k_{t-1}^{1:5} \in [N]^5} \mathsf{T}_{t-1,b}^{(2)}(k_{t-1}^1, k_{t-1}^2, k_{t-1}^3) \tilde{\mathcal{T}}_{t-1}^b(k_{t-1}^4, k_{t-1}^5) \sum_{k_t^{1:5} \in S_5} \mathbb{E} \left[ \prod_{n=1}^5 \beta_t^{\text{BS}}(k_t^n, k_{t-1}^n) \middle| \mathcal{G}_{t-1}^N \right]. \end{aligned}$$

Proceeding similarly as in Proposition A.6 and Lemma A.5, it can be shown that

$$\sum_{k_t^{1:5} \in S_5} \mathbb{E} \left[ \prod_{n=1}^5 \beta_t^{\text{BS}}(k_t^n, k_{t-1}^n) \middle| \mathcal{G}_{t-1}^N \right] \leq \frac{N^5 G_\infty^5}{\Omega_{t-1}^5} + \frac{2N^4 G_\infty^4}{\Omega_{t-1}^4} + \frac{N^3 G_\infty^3}{\Omega_{t-1}^3}.$$

Finally, by noting that

$$\sum_{k_{t-1}^{1:5} \in [N]^5} \mathsf{T}_{t-1,b}^{(2)}(k_{t-1}^1, k_{t-1}^2, k_{t-1}^3) \tilde{\mathcal{T}}_{t-1}^b(k_{t-1}^4, k_{t-1}^5) \leq 4 \sum_{k_{t-1}^{1:6} \in [N]^6} \prod_{\ell=1}^3 \tilde{\mathcal{T}}_{t-1}^b(k_{t-1}^{2\ell-1}, k_{t-1}^{2\ell})$$

we get

$$\begin{aligned} &\mathbb{E}[F_{1,b}^N F_{3,b}^N | \mathcal{G}_{t-1}^N] \\ &\leq \left( \frac{N^5 G_\infty^5}{\Omega_{t-1}^5} + \frac{2N^4 G_\infty^4}{\Omega_{t-1}^4} + \frac{N^3 G_\infty^3}{\Omega_{t-1}^3} \right) \sum_{k_{t-1}^{1:5} \in [N]^5} \mathsf{T}_{t-1,b}^{(2)}(k_{t-1}^1, k_{t-1}^2, k_{t-1}^3) \tilde{\mathcal{T}}_{t-1}^b(k_{t-1}^4, k_{t-1}^5) \\ &\leq 4 \left( \frac{N^5 G_\infty^5}{\Omega_{t-1}^5} + \frac{2N^4 G_\infty^4}{\Omega_{t-1}^4} + \frac{N^3 G_\infty^3}{\Omega_{t-1}^3} \right) \sum_{k_{t-1}^{1:6} \in [N]^6} \prod_{\ell=1}^3 \tilde{\mathcal{T}}_{t-1}^b(k_{t-1}^{2\ell-1}, k_{t-1}^{2\ell}). \end{aligned} \quad (\text{A1.54})$$



Consider now the case  $\mathbb{E}[F_{1,b}^N F_{2,b}^N | \mathcal{G}_{t-1}^N]$ . In the same way as for the previous case, we obtain

$$\mathbb{E}[F_{1,b}^N F_{2,b}^N | \mathcal{G}_{t-1}^N] \leq 2 \left( \frac{N^4 G_\infty^4}{\Omega_{t-1}^4} + \frac{2N^3 G_\infty^3}{\Omega_{t-1}^3} + \frac{N^2 G_\infty^2}{\Omega_{t-1}^2} \right) \sum_{k_{t-1}^{1:6} \in [N]^6} \prod_{\ell=1}^3 \tilde{\tau}_{t-1}^b(k_{t-1}^{2\ell-1}, k_{t-1}^{2\ell}). \quad (\text{A1.55})$$

Thus, combining (A1.55), (A1.54) and (A1.53), we obtain

$$\begin{aligned} & \mathbb{E} \left[ \mathcal{C}_{N,b,t}^3 \sum_{k_t^{1:6} \in [N]^6} \sum_{i^{1:3} \in [M]^3} \mathbb{1}_{i^1=i^2, i^3 \notin \{i^1, i^2\}} \prod_{\ell=1}^3 \tilde{\tau}_{t-1}^b(J_{k_t^{2\ell-1}, t-1}^{i^\ell}, J_{k_t^{2\ell}, t-1}^{i^\ell}) \right] \\ & \lesssim \mathbb{E} \left[ \left( \sum_{p=2}^6 \frac{\Omega_{t-1}^6 N^p G_\infty^p}{\Omega_{t-1}^p N^3 (N-1)^3} \right) \tilde{\mathcal{Q}}_{b,t-1}^{N,M}(\mathbf{1})^3 \right] \lesssim \|\tilde{\mathcal{Q}}_{b,t-1}^{N,M}(\mathbf{1})\|_3^3. \end{aligned}$$

The other triplets  $(i^1, i^2, i^3)$  for which  $\text{Card}(\{i^1, i^2, i^3\}) = 2$  are handled similarly and are bounded by  $\|\tilde{\mathcal{Q}}_{b,t-1}^{N,M}(\mathbf{1})\|_3^3$  up to some multiplicative constant independent of  $N$ . We finally obtain

$$\mathbb{E} \left[ \frac{\mathcal{C}_{N,b,t}^3}{M^3} \sum_{k_t^{1:6} \in [N]^6} \sum_{i^{1:3} \in [M]^3} \mathbb{1}_{\text{Card}(\{i^1, i^2, i^3\})=2} \prod_{\ell=1}^3 \tilde{\tau}_{t-1}^b(J_{k_t^{2\ell-1}, t-1}^{i^\ell}, J_{k_t^{2\ell}, t-1}^{i^\ell}) \right] \lesssim \|\tilde{\mathcal{Q}}_{b,t-1}^{N,M}(\mathbf{1})\|_3^3. \quad (\text{A1.56})$$

It now remains to treat the case  $\text{Card}(\{i^1, i^2, i^3\}) = 1$ . Let  $(k^1, \dots, k^d) \in [N]^d$ . Denote by  $\text{Pos}_d(k^{1:d})$  the set of elements in  $[N]^d$  with positions of equal elements similar to those of the equal elements in  $k^{1:d}$ . For example,

$$\begin{aligned} \text{Pos}_3((2, 1, 1)) &= \{(j, i, i) \mid (i, j) \in [N]^2, i \neq j\}, \\ \text{Pos}_3((1, 1, 2)) &= \{(i, i, j) \mid (i, j) \in [N]^2, i \neq j\}, \\ \text{Pos}_3((1, 2, 3)) &= \{(i, j, k) \mid (i, j, k) \in [N]^3, \text{Card}(\{i, j, k\}) = 3\}. \end{aligned}$$

Let  $p \in [2 : 6]$  and  $(k_t^1, \dots, k_t^6) \in \mathcal{I}_0^3 \cap \mathcal{S}_3^p$ . Without loss of generality, assume that the first  $p$  elements of  $k_t^{1:6}$  are all different.

$$\begin{aligned} & \mathbb{E} \left[ \prod_{\ell=1}^3 \tilde{\tau}_{t-1}^b(J_{k_t^{2\ell-1}, t-1}^{i^\ell}, J_{k_t^{2\ell}, t-1}^{i^\ell}) \middle| \mathcal{G}_{t-1}^N \right] \\ &= \sum_{k_{t-1}^{1:6} \in \text{Pos}_6(k_t^{1:6})} \prod_{\ell=1}^3 \tilde{\tau}_{t-1}^b(k_{t-1}^{2\ell-1}, k_{t-1}^{2\ell}) \mathbb{E} \left[ \prod_{n=1}^p \beta_t^{\text{BS}}(k_t^n, k_{t-1}^n) \middle| \mathcal{G}_{t-1}^N \right] \\ &\leq \frac{G_\infty^p}{\Omega_{t-1}^p} \sum_{k_{t-1}^{1:6} \in \text{Pos}_6(k_t^{1:6})} \prod_{\ell=1}^3 \tilde{\tau}_{t-1}^b(k_{t-1}^{2\ell-1}, k_{t-1}^{2\ell}) \\ &\leq \frac{G_\infty^p}{\Omega_{t-1}^p} \sum_{k_{t-1}^{1:6} \in [N]^6} \prod_{\ell=1}^3 \tilde{\tau}_{t-1}^b(k_{t-1}^{2\ell-1}, k_{t-1}^{2\ell}). \end{aligned}$$

Consequently, by Lemma A.5,

$$\begin{aligned}
& \sum_{k_t^{1:6} \in \mathcal{I}_0^3 \cap \mathcal{S}_3^p} \mathbb{E} \left[ \prod_{\ell=1}^3 \tilde{\mathcal{T}}_{t-1}^b(J_{k_t^{2\ell-1}, t-1}^{i^\ell}, J_{k_t^{2\ell}, t-1}^{i^\ell}) \middle| \mathcal{G}_{t-1}^N \right] \\
& \leq \sum_{k_t^{1:6} \in \mathcal{I}_0^3 \cap \mathcal{S}_3^p} \frac{G_\infty^p}{\Omega_{t-1}^p} \sum_{k_{t-1}^{1:6} \in [N]^6} \prod_{\ell=1}^3 \tilde{\mathcal{T}}_{t-1}^b(k_{t-1}^{2\ell-1}, k_t^{2\ell}) \\
& \lesssim \frac{N^p G_\infty^p}{\Omega_{t-1}^p} \sum_{k_{t-1}^{1:6} \in [N]^6} \prod_{\ell=1}^3 \tilde{\mathcal{T}}_{t-1}^b(k_{t-1}^{2\ell-1}, k_t^{2\ell}),
\end{aligned}$$

and,

$$\begin{aligned}
& \mathbb{E} \left[ \frac{\mathcal{C}_{N,b,t}^3}{M^3} \sum_{k_t^{1:6} \in [N]^6} \sum_{i^{1:3} \in [M]^3} \mathbb{1}_{\text{Card}(\{i^1, i^2, i^3\})=1} \prod_{\ell=1}^3 \tilde{\mathcal{T}}_{t-1}^b(J_{k_t^{2\ell-1}, t-1}^{i^\ell}, J_{k_t^{2\ell}, t-1}^{i^\ell}) \right] \\
& \lesssim \mathbb{E} \left[ \left( \sum_{p=2}^6 \frac{\Omega_{t-1}^6 N^p G_\infty^p}{\Omega_{t-1}^p N^3 (N-1)^3} \right) \tilde{\mathcal{Q}}_{b,t-1}^{N,M}(\mathbf{1})^3 \right] \lesssim \|\tilde{\mathcal{Q}}_{b,t-1}^{N,M}(\mathbf{1})\|_3^3. \quad (\text{A1.57})
\end{aligned}$$

Finally, combining (A1.51), (A1.56) and (A1.57) we get

$$\|\tilde{\mathcal{Q}}_{b,t}^{N,M}(\mathbf{1})\|_3^3 = \mathbb{E} \left[ \frac{\mathcal{C}_{N,b,t}^3}{M^3} \sum_{k_t^{1:6} \in [N]^6} \sum_{i^{1:3} \in [M]^3} \prod_{\ell=1}^3 \tilde{\mathcal{T}}_{t-1}^b(J_{k_t^{2\ell-1}, t-1}^{i^\ell}, J_{k_t^{2\ell}, t-1}^{i^\ell}) \right] \lesssim \|\tilde{\mathcal{Q}}_{b,t-1}^{N,M}(\mathbf{1})\|_3^3,$$

and hence  $\sup_{N \geq 6} \|\tilde{\mathcal{Q}}_{b,t}^{N,M}(\mathbf{1})\|_3 < \infty$  by the induction hypothesis. This ends the proof for the case  $b_t = 0$ .

If  $M = 2$ , then (A1.51) is equal to 0 and (A1.56), (A1.57) remain the same. The result then follows. If  $N < 6$  then it suffices to truncate the sums over  $p$  to obtain the result.  $\square$

## APPENDIX B. FURTHER ALGORITHMIC DETAILS

**B.1. Variance estimators for the predictor and filter.** The asymptotic variances of the predictor and filter (3.8)-(3.9) can be expressed using  $\mathcal{V}_{\gamma,t}^\infty$ . Indeed,

$$\frac{\mathcal{V}_{\gamma,t}^\infty(h - \eta_t(h))}{\gamma_t(\mathbf{1})^2} = \sum_{s=0}^t \left\{ \frac{\gamma_s(\mathbf{1}) \gamma_s(\mathbf{Q}_{s:t}[h - \eta_t(h)]^2)}{\gamma_t(\mathbf{1})^2} - \eta_t(h - \eta_t(h))^2 \right\} = \mathcal{V}_{\eta,t}^\infty(h),$$

and using that

$$\frac{\gamma_t(g_t\{h - \phi_t(h)\})}{\gamma_{t+1}(\mathbf{1})} = \frac{\gamma_t(g_th)}{\gamma_{t+1}(\mathbf{1})} - \phi_t(h) = 0,$$

we get

$$\mathcal{V}_{\phi,t}^\infty(h) = \frac{\mathcal{V}_{\gamma,t}^\infty(g_t\{h - \phi_t(h)\})}{\gamma_{t+1}(\mathbf{1})^2}. \quad (\text{A2.1})$$

Then, replacing  $\gamma_t(h)$  and  $\phi_t(h)$  by their empirical approximations  $\gamma_t^N(h)$  and  $\phi_t^N(h)$ , we obtain

$$\mathcal{V}_{\eta,t}^{N,BS}(h) := \frac{-N^t}{(N-1)^{t+1}} \sum_{i,j \in [N]^2} \mathcal{T}_t^0(i,j) \{h(\xi_t^i) - \eta_t^N(h)\} \{h(\xi_t^j) - \eta_t^N(h)\}, \quad (\text{A2.2})$$

$$\mathcal{V}_{\phi,t}^{N,BS}(h) := \frac{-N^{t+2}}{(N-1)^{t+1}} \sum_{i,j \in [N]^2} \mathcal{W}_t^i \mathcal{W}_t^j \mathcal{T}_t^0(i,j) \{h(\xi_t^i) - \phi_t^N(h)\} \{h(\xi_t^j) - \phi_t^N(h)\}. \quad (\text{A2.3})$$

As a consequence of Theorem 4.5, these estimators are also weakly consistent.

**Corollary B.1.** *Let (A1 : 4) hold. For any  $h \in \mathbb{F}_b(\mathcal{X})$ ,  $\mathcal{V}_{\eta,t}^{N,BS}(h) \xrightarrow{\mathbb{P}} \mathcal{V}_{\eta,t}^\infty(h)$  and  $\mathcal{V}_{\phi,t}^{N,BS}(h) \xrightarrow{\mathbb{P}} \mathcal{V}_{\phi,t}^\infty(h)$ .*

*Proof.* It suffices to note that  $\mathcal{Q}_{b,t}^{N,BS}$  and  $\mathcal{Q}_{b,t}$  are bilinear, that  $\eta_t^N(h) \xrightarrow{\mathbb{P}} \eta_t(h)$  and to apply Theorem 4.4 again:

$$\begin{aligned} & \mathcal{Q}_{b,t}^{N,BS}(\{h - \eta_t^N(h)\}^{\otimes 2}) \\ &= \mathcal{Q}_{b,t}^{N,BS}(h^{\otimes 2}) - \eta_t^N(h) \mathcal{Q}_{b,t}^{N,BS}(h \otimes \mathbf{1}) - \eta_t^N(h) \mathcal{Q}_{b,t}^{N,BS}(\mathbf{1} \otimes h) + \eta_t^N(h)^2 \mathcal{Q}_{b,t}^{N,BS}(\mathbf{1}) \\ &\xrightarrow{\mathbb{P}} \mathcal{Q}_{b,t}(\{h - \eta_t(h)\}^{\otimes 2}). \end{aligned}$$

Hence,  $\mathcal{V}_{\gamma,t}^{N,BS}(h - \eta_t^N(h)) \xrightarrow{\mathbb{P}} \mathcal{V}_{\gamma,t}^\infty(h - \eta_t(h))$  and using the fact that  $\gamma_t^N(\mathbf{1})^2 \xrightarrow{\mathbb{P}} \gamma_t(\mathbf{1})^2$  we get the consistency for the predictive measures. The remaining limit is a straightforward application.  $\square$

---

**Algorithm 3** Variance estimator for predictive measures

---

**Require:**  $\omega_t^{1:N}, \xi_t^{1:N}, \xi_{t+1}^{1:N}$  and  $\mathcal{T}_t^0$

Compute  $\beta_{t+1}^{BS}$

**if** *PaRIS* **then**

**for**  $k \in [1 : N]$  **do**

    Sample  $J_{k,t}^{1:M} \stackrel{\text{iid}}{\sim} \beta_{t+1}^{BS}(k, \cdot)$

**end for**

**for**  $(k, \ell) \in [1 : N]^2$  **do**

    Set  $\mathcal{T}_{t+1}^0(k, \ell) = \mathbb{1}_{k \neq \ell} \sum_{i=1}^M \mathcal{T}_t^0(J_{k,t}^i, J_{\ell,t}^i) / M$

**end for**

**else**

  Compute  $\overline{\mathcal{T}}_{t+1}^0 = \beta_{t+1}^{BS} \mathcal{T}_t^0 \beta_{t+1}^{BS'}$ .

  Set  $\mathcal{T}_{t+1}^0 = \overline{\mathcal{T}}_{t+1}^0 - \text{Diag}(\overline{\mathcal{T}}_{t+1}^0)$ .

**end if**

Compute  $\mathcal{Q} = \mathcal{T}_{t+1}^0 \odot [\{h(\xi_{t+1}^{1:N}) - \eta_{t+1}^N(h)\} \{h(\xi_{t+1}^{1:N}) - \eta_{t+1}^N(h)\}^\top]$ .

$\triangleright h$  is applied elementwise

**return**  $-N^t / (N-1)^{t+1} \sum_{i,j \in [N]^2} \mathcal{Q}_{i,j}, \quad \mathcal{T}_{t+1}^0$ .

---

**B.2. GT term by term estimator of the asymptotic variance.** In this section we derive the GT counterpart of the term by term estimator (4.17). Define for all  $t > 0$

$$\mathcal{T}_{b,t}^{\text{GT}}(K_t^1, K_t^2) := \mathbb{E}_{\text{GT}}[\mathbb{I}_{b,t}(K_{0:t}^1, K_{0:t}^2) | \mathcal{F}_t^N, K_t^1, K_t^2], \quad (\text{A2.4})$$

and  $\mathcal{T}_{b,0}^{\text{GT}}(K_0^1, K_0^2) := \mathbb{1}_{K_0^1 \neq K_0^2, b_0=0} + \mathbb{1}_{K_0^1 = K_0^2, b_0=1}$ . By the tower property and the definition of  $\mathbb{E}_{\text{GT}}[\cdot | \mathcal{F}_{t-1}^N]$ , for all  $(k, \ell) \in [N]^2$  and  $t > 0$ , if  $b_t = 0$ ,

$$\mathcal{T}_{b,t}^{\text{GT}}(k, \ell) = \mathbb{1}_{k \neq \ell} \sum_{i,j \in [N]^2} \mathbb{1}_{A_t^k=i, A_t^\ell=j} \mathcal{T}_{b,t-1}^{\text{GT}}(i, j) = \mathbb{1}_{k \neq \ell} \mathcal{T}_{b,t-1}^{\text{GT}}(A_{t-1}^k, A_{t-1}^\ell),$$

and if  $b_t = 1$ ,

$$\mathcal{T}_{b,t}^{\text{GT}}(k, \ell) = \mathbb{1}_{k=\ell} \sum_{i,j \in [N]^2} \mathbb{1}_{A_{t-1}^k=i} \mathcal{W}_{t-1}^j \mathcal{T}_{b,t-1}^{\text{GT}}(i, j) = \mathbb{1}_{k=\ell} \sum_{i=1}^N \mathcal{W}_{t-1}^i \mathcal{T}_{b,t-1}^{\text{GT}}(A_{t-1}^k, i).$$

Similarly to BS, we have by the tower property

$$\mathcal{Q}_{b,t}^{N,\text{BS}}(h) = \prod_{s=0}^t N^{b_s} \left( \frac{N}{N-1} \right)^{1-b_s} \frac{\gamma_t^N(\mathbf{1})^2}{N^2} \sum_{k,\ell \in [N]^2} \mathcal{T}_{b,t}^{\text{GT}}(k, \ell) h(\xi_t^k, \xi_t^\ell), \quad (\text{A2.5})$$

and the term by term estimator is thus

$$\bar{\mathcal{V}}_{\gamma,t}^{N,\text{GT}}(h) = \frac{N^{t-1} \gamma_t^N(\mathbf{1})^2}{(N-1)^t} \sum_{k,\ell \in [N]^2} \left\{ S_t^{\text{GT}}(k, \ell) - \frac{t+1}{N-1} \mathcal{T}_{0,t}^{\text{GT}}(k, \ell) \right\} h(\xi_t^k) h(\xi_t^\ell), \quad (\text{A2.6})$$

where  $S_t^{\text{GT}}$  is such that for all  $(k, \ell) \in [N]^2$ ,

$$S_t^{\text{GT}}(k, \ell) = \sum_{s=0}^t \mathcal{T}_{e_s,t}^{\text{GT}}(k, \ell) = \mathbb{1}_{k=\ell} \sum_{i=1}^N \mathcal{W}_{t-1}^i \mathcal{T}_{0,t-1}^{\text{GT}}(A_{t-1}^k, i) + \mathbb{1}_{k \neq \ell} S_{t-1}^{\text{GT}}(A_{t-1}^k, A_{t-1}^\ell),$$

which shows that (A2.6) is also updated online in a rather simple way by propagating the matrices  $S_t^{\text{GT}}$  and  $\mathcal{T}_{0,t}^{\text{GT}}$ .

## APPENDIX C. TECHNICAL RESULTS

**Theorem C.1 (Generalized dominated convergence theorem).** *Let  $(f_N)_{N \in \mathbb{N}}$  be a sequence of  $\mathcal{X}$ -measurable functions and  $(g_N)_{N \in \mathbb{N}}$  a sequence of non-negative  $\mathcal{X}$ -measurable functions. Assume that the following assumptions hold.*

- (i) *There exists  $C > 0$  such that  $|f_N(x)| \leq C g_N(x)$  for all  $N \in \mathbb{N}$  and  $x \in \mathcal{X}$ .*
- (ii)  *$(g_N)_{N \in \mathbb{N}}$  converges pointwise to  $g$  and  $\lim_{N \rightarrow \infty} \int g_N d\nu = \int g d\nu < \infty$ .*
- (iii)  *$(f_N)_{N \in \mathbb{N}}$  converges pointwise to  $f$ .*

*Then,  $f$  is  $\nu$ -integrable and  $\lim_{N \rightarrow \infty} \int f_N d\nu = \int f d\nu$ .*

*Proof.* The proof can be found in [36]. □

Theorem C.2 and Lemma C.3 are borrowed from [33] and [15] respectively.

**Theorem C.2.** Assume that (A2 : 4) hold. Then, for all  $s \in \mathbb{N}$ ,  $h_s \in \mathbb{F}_b(\mathcal{X}^{\otimes s+1})$  and  $(f_s, \tilde{f}_s) \in \mathbb{F}_b(\mathcal{X})^2$ , there exist constants  $(C_s, \tilde{C}_s) \in (\mathbb{R}_+^*)^2$ , depending on  $h_s, f_s$ , and  $\tilde{f}_s$ , such that for all  $N \in \mathbb{N}^*$  and all  $\varepsilon \in \mathbb{R}_+^*$ ,

$$\mathbb{P} \left( \left| N^{-1} \sum_{i=1}^N \omega_s^i \{ (\mathbf{T}_s^N[h_s]f_s)(\xi_s^i) + \tilde{f}_s(\xi_s^i) \} - \eta_s(\mathbf{T}_s[h_s]f_s + \tilde{f}_s) \right| \geq \varepsilon \right) \leq C_s \exp(-\tilde{C}_s N \varepsilon^2), \quad (\text{A3.1})$$

$$\mathbb{P} \left( \left| \sum_{i=1}^N \mathcal{W}_s^i \{ (\mathbf{T}_s^N[h_s]f_s)(\xi_s^i) + \tilde{f}_s(\xi_s^i) \} - \phi_s(\mathbf{T}_s[h_s]f_s + \tilde{f}_s) \right| \geq \varepsilon \right) \leq C_s \exp(-\tilde{C}_s N \varepsilon^2). \quad (\text{A3.2})$$

**Lemma C.3.** Assume that  $a_N, b_N$  and  $b$  are random variables defined on the same probability space such that there exist positive constants  $\beta, B_1, C_1, B_2, C_2$  and  $M$  satisfying the following assumptions.

- $|a_N/b_N| \leq M$ ,  $\mathbb{P}$ -a.s. and  $b \geq \beta$ .
- For all  $\varepsilon > 0$  and all  $N \geq 1$ ,  $\mathbb{P}(|b_N - b| > \varepsilon) \leq B_1 \exp(-C_1 N \varepsilon^2)$ .
- For all  $\varepsilon > 0$  and all  $N \geq 1$ ,  $\mathbb{P}(|a_N| > \varepsilon) \leq B_2 \exp(-C_2 N \varepsilon^2)$ .

Then, there exist two positive constants  $B_3, C_3$  such that

$$\mathbb{P} \left( \left| \frac{a_N}{b_N} \right| > \varepsilon \right) \leq B_3 \exp(-C_3 N \varepsilon^2).$$

#### APPENDIX D. ASYMPTOTIC VARIANCE OF THE JOINT PREDICTIVE DISTRIBUTION

In this section we provide some intuition on (3.7). Let  $h \in \mathbb{F}_b(\mathcal{X})$ . By the law of total variance,

$$\mathbb{V}[\gamma_{t+1}^N(h)] = \mathbb{V}[\mathbb{E}[\gamma_{t+1}^N(h) | \mathcal{F}_t^N]] + \mathbb{E}[\mathbb{V}[\gamma_{t+1}^N(h) | \mathcal{F}_t^N]]. \quad (\text{A4.1})$$

As  $\gamma_{t+1}^N(\mathbf{1})$  is  $\mathcal{F}_t^N$ -measurable and the particles at time  $t+1$  are i.i.d conditionally on  $\mathcal{F}_t^N$ , we have that

$$\begin{aligned} \mathbb{E}[\gamma_{t+1}^N(h) | \mathcal{F}_t^N] &= \gamma_{t+1}^N(\mathbf{1}) \sum_{i=1}^N \frac{\omega_t^i}{\Omega_t} \mathbf{M}_{t+1}[h](\xi_t^i) \\ &= \gamma_t^N(\mathbf{1}) N^{-1} \Omega_t \sum_{i=1}^N \frac{\omega_t^i}{\Omega_t} \mathbf{M}_{t+1}[h](\xi_t^i) = \gamma_t^N(\mathbf{Q}_{t+1}[h]). \end{aligned} \quad (\text{A4.2})$$

On the other hand,

$$\mathbb{V}[\gamma_{t+1}^N(h) | \mathcal{F}_t^N] = \gamma_{t+1}^N(\mathbf{1})^2 \mathbb{V} \left[ \frac{1}{N} \sum_{i=1}^N h(\xi_{t+1}^i) \middle| \mathcal{F}_t^N \right] = N^{-1} \gamma_{t+1}^N(\mathbf{1})^2 \mathbb{V}_{\phi_t^N \mathbf{M}_{t+1}}[h(\xi_{t+1})]$$

where

$$\mathbb{V}_{\phi_t^N \mathbf{M}_{t+1}}[h(\xi_{t+1})] = \phi_t^N \mathbf{M}_{t+1}(\{h - \phi_t^N \mathbf{M}_{t+1}(h)\}^2).$$

Therefore,

$$\begin{aligned} \mathbb{V}[\gamma_{t+1}^N(h) | \mathcal{F}_t^N] &= N^{-2} \gamma_t^N(\mathbf{1})^2 \Omega_t \eta_t^N(\mathbf{Q}_{t+1}[\{h - \phi_t^N \mathbf{M}_{t+1}(h)\}^2]) \\ &= N^{-1} \gamma_{t+1}^N(\mathbf{1}) \gamma_t^N(\mathbf{Q}_{t+1}[\{h - \phi_t^N \mathbf{M}_{t+1}(h)\}^2]). \end{aligned} \quad (\text{A4.3})$$

Replacing (A4.2)-(A4.3) in (A4.1), we get the recursive formula

$$\begin{aligned} N\mathbb{V}[\gamma_{t+1}^N(h)] &= N\mathbb{V}[\gamma_t^N(\mathbf{Q}_{t+1}[h])] + \mathbb{E}[\gamma_{t+1}^N(\mathbf{1})\gamma_t^N(\mathbf{Q}_{t+1}[\{h - \phi_t^N M_{t+1}(h)\}^2])] \\ &= M_0(\mathbf{Q}_{1:t+1}[h]^2) - \gamma_{t+1}(h)^2 \\ &\quad + \sum_{s=1}^{t+1} \mathbb{E}[\gamma_s^N(\mathbf{1})\gamma_{s-1}^N(\mathbf{Q}_s[\{\mathbf{Q}_{s+1:t+1}[h] - \phi_{s-1}^N(M_s[\mathbf{Q}_{s+1:t+1}[h]])\}^2])] . \end{aligned}$$

With multiple applications of (3.5) in the main paper, we get that

$$\begin{aligned} &\gamma_s^N(\mathbf{1})\gamma_{s-1}^N(\mathbf{Q}_s[\{\mathbf{Q}_{s+1:t+1}[h] - \phi_{s-1}^N(M_s[\mathbf{Q}_{s+1:t+1}[h]])\}^2]) \\ &\xrightarrow{a.s.} \gamma_s(\mathbf{1})\gamma_{s-1}(\mathbf{Q}_s[\{\mathbf{Q}_{s+1:t+1}[h] - \phi_{s-1}(M_s[\mathbf{Q}_{s+1:t+1}[h]])\}^2]) , \quad (\text{A4.4}) \end{aligned}$$

and, using that  $\gamma_s(\mathbf{1})\phi_{s-1}(M_s[h]) = \gamma_s(h)$  and  $\gamma_s(\mathbf{Q}_{s+1:t+1}[h]) = \gamma_{t+1}(h)$  for all  $h$  we get

$$\begin{aligned} &\gamma_s(\mathbf{1})\gamma_{s-1}(\mathbf{Q}_s[\{\mathbf{Q}_{s+1:t+1}[h] - \phi_{s-1}(M_s[\mathbf{Q}_{s+1:t+1}[h]])\}^2]) \\ &= \gamma_s(\mathbf{1})\gamma_s(\{\mathbf{Q}_{s+1:t+1}[h] - \phi_{s-1}(M_s[\mathbf{Q}_{s+1:t+1}[h]])\}^2) \\ &= \gamma_s(\mathbf{1})\gamma_s[\mathbf{Q}_{s+1:t+1}(h)^2] - \gamma_{t+1}(h)^2 . \end{aligned}$$

Finally, by (A2) and the boundedness of  $h$ ,  $|\gamma_s^N(\mathbf{1})| \leq G_\infty^s$ ,  $|\mathbf{Q}_{s+1:t}[h]| \leq G_\infty^{t-s}|h|_\infty$ , thus

$$|\gamma_s^N(\mathbf{1})\gamma_{s-1}^N(\mathbf{Q}_s[\{\mathbf{Q}_{s+1:t+1}[h] - \phi_{s-1}^N(M_s[\mathbf{Q}_{s+1:t+1}[h]])\}^2])| \leq 4G_\infty^{2t}|h|_\infty ,$$

and by the dominated convergence theorem, for any  $s \in [0 : t]$ ,

$$\begin{aligned} &\lim_{N \rightarrow \infty} \mathbb{E}[\gamma_s^N(\mathbf{1})\gamma_{s-1}^N(\mathbf{Q}_s[\{\mathbf{Q}_{s+1:t+1}[h] - \phi_{s-1}^N(M_s[\mathbf{Q}_{s+1:t+1}[h]])\}^2])] \\ &= \gamma_s(\mathbf{1})\gamma_s[\mathbf{Q}_{s+1:t+1}(h)^2] - \gamma_{t+1}(h)^2 . \quad (\text{A4.5}) \end{aligned}$$

and  $N\mathbb{V}[\gamma_{t+1}^N(h)] \rightarrow \sum_{s=0}^{t+1} \{\gamma_s(\mathbf{1})\gamma_s[\mathbf{Q}_{s+1:t+1}(h)^2] - \gamma_{t+1}(h)^2\}$ . As argued in Section 4.3 of the main paper,  $\mathbb{E}[N(\gamma_{t+1}^N(h) - \gamma_{t+1}(h))^2]$  converges to the asymptotic variance when  $N$  goes to infinity, and since  $\gamma_{t+1}^N(h)$  is an unbiased estimator of  $\gamma_{t+1}(h)$ ,  $N\mathbb{V}[\gamma_{t+1}^N(h)] = \mathbb{E}[N(\gamma_{t+1}^N(h) - \gamma_{t+1}(h))^2]$ . Therefore,

$$\lim_{N \rightarrow \infty} \mathbb{E}[N(\gamma_{t+1}^N(h) - \gamma_{t+1}(h))^2] = \sum_{s=0}^{t+1} \{\gamma_s(\mathbf{1})\gamma_s[\mathbf{Q}_{s+1:t+1}(h)^2] - \gamma_{t+1}(h)^2\} ,$$

which ends the proof.