A pseudo-marginal sequential Monte Carlo online smoothing algorithm

Pierre Gloaguen, Sylvain Le Corff, Jimmy Olsson

A pseudo-marginal sequential Monte Carlo online smoothing algorithm

Pierre Gloaguen[1], Sylvain Le Corff[2] and Jimmy Olsson[3]

[1] *AgroParisTech, Paris, France. e-mail:* pierre.gloaguem@agroparistech.fr

[2] *Télécom SudParis, Paris, France. e-mail:* sylvain_lecorff@telecom-sudparis.eu

[3] *KTH Royal Institute of Technology, Stockholm, Sweden. e-mail:* jimmyol@kth.se

Abstract: We consider online computation of expectations of additive state functionals under general path probability measures proportional to products of unnormalised transition densities. These transition densities are assumed to be intractable but possible to estimate, with or without bias. Using pseudo-marginalisation techniques we are able to extend the particle-based, rapid incremental smoother (PaRIS) algorithm proposed in [J. Olsson and J. Westerborn. Efficient particle-based online smoothing in general hidden Markov models: The PaRIS algorithm. *Bernoulli*, 23(3):1951–1996, 2017] to this setting. The resulting algorithm, which has a linear complexity in the number of particles and constant memory requirements, applies to a wide range of challenging path-space Monte Carlo problems, including smoothing in partially observed diffusion processes and models with intractable likelihood. The algorithm is furnished with several theoretical results, including a central limit theorem, establishing its convergence and numerical stability. Moreover, under strong mixing assumptions we establish a novel $\mathcal{O}(n\varepsilon)$ bound on the asymptotic bias of the algorithm, where n is the path length and ε controls the bias of the density estimators.

Keywords and phrases: central limit theorem, exponential concentration, partially observed diffusions, particle smoothing, pseudo-marginal methods, sequential Monte Carlo methods.

1. Introduction

Let $(\mathsf{X}_n, \mathcal{X}_n)_{n \in \mathbb{N}}$ be a sequence of general state spaces and let, for all $n \in \mathbb{N}$, $\mathbf{L}_n : \mathsf{X}_n \times \mathcal{X}_{n+1} \to \mathbb{R}_{\geq 0}$ be bounded kernels in the sense that $\sup_{x \in \mathsf{X}_n} \mathbf{L}_n(x, \mathsf{X}_{n+1}) < \infty$. We will assume a dominated model where each kernel $\mathbf{L}_n$ has a kernel density ℓ_n with respect to some σ-finite reference measure μ_{n+1} on $\mathcal{X}_{n+1}$. Finally, let χ be some bounded measure on $\mathcal{X}_0$. In the following, we denote state-space product sets and σ-fields by $\mathsf{X}^n \coloneqq \mathsf{X}_0 \times \cdots \times \mathsf{X}_n$ and $\mathcal{X}^n \coloneqq \mathcal{X}_0 \otimes \cdots \otimes \mathcal{X}_n$, respectively, and consider probability measures

$$\phi_{0:n}(dx_{0:n}) \propto \chi(dx_0) \prod_{m=0}^{n-1} \mathbf{L}_m(x_m, dx_{m+1}), \quad n \in \mathbb{N}, \tag{1.1}$$

on these product spaces.[1] Given a sequence $(\tilde{h}_n)_{n \in \mathbb{N}}$ of measurable functions $\tilde{h}_n : \mathsf{X}_n \times \mathsf{X}_{n+1} \to \mathbb{R}$, the aim of the present paper is the online approximation

---

[1] We will always use the standard convention $\prod_{\varnothing} = 1$, implying that $\phi_0 \propto \chi$.

of expectations of *additive functionals*

$$h_n : \mathsf{X}^n \ni x_{0:n} \mapsto \sum_{m=0}^{n-1} \tilde{h}_m(x_m, x_{m+1}) \tag{1.2}$$

under the distribution flow $(\phi_{0:n})_{n \in \mathbb{N}}$ using *sequential Monte Carlo* (SMC) methods.

The generality of the model (1.1) is striking. In the special case where each $\mathbf{L}_n$ can be decomposed as $\mathbf{L}_n(x_n, dx_{n+1}) = g_n(x_n)\,\mathbf{Q}_n(x_n, dx_{n+1})$ for some Markov kernel $\mathbf{Q}_n$ and some nonnegative potential function g_n, (1.1) yields the *Feynman-Kac path models* [9], which are applied in a large variety of scientific and engineering disciplines, including statistics, physics, biology, and signal processing. In a *hidden Markov model* (HMM) (see, *e.g.*, [7]), a Markov chain $(X_n)_{n \in \mathbb{N}}$ with kernels $(\mathbf{Q}_n)_{n \in \mathbb{N}}$ and initial distribution χ is only partially observed through a sequence $(Y_n)_{n \in \mathbb{N}}$ of observations being conditionally independent given the Markov states. In that case, g_n plays the role of the likelihood of the state X_n given the corresponding observation Y_n, and $\phi_{0:n}$ describes the *joint-smoothing distribution*, *i.e.*, the joint posterior of the hidden states $X_0, \ldots, X_n$ given corresponding observations (see Example 1 for details). We will adopt this terminology throughout the present paper and refer to the distributions defined in (1.1) as 'smoothing distributions'; the problem of computing expectations of functionals of type (1.2) under these distributions will be referred to as 'additive smoothing'. General state-space HMMs are prevalent in time-series and sequential-data analysis and are used extensively in, *e.g.*, movement ecology [26], energy-consumption modeling [5], genomics [37], target tracking [35], enhancement and segmentation of speech and audio signals [32]; see also [34, 39] and the numerous references therein. Operating on models of this sort, online additive smoothing is instrumental for, *e.g.*,

– *path reconstruction*, *i.e.*, the estimation of hidden states given observations. Especially in *fixed-point smoothing*, where interest is in computing the expectations of $h(X_m)$ conditionally to $Y_0, \ldots, Y_n$ for some given m and test function h as n tends to infinity, a problem that can be cast into our framework by letting, in (1.2), $\tilde{h}_m(x_m, x_{m+1}) = h(x_m)$ and $\tilde{h}_\ell \equiv 0$ for all $\ell \neq m$.
– *parameter inference*, where additive smoothing is a key ingredient in the computation of log-likelihood gradients (*score functions*) via Fisher's identity or the intermediate quantity of the *expectation-maximisation* (EM) *algorithm*; see, *e.g.*, [7, Chapter 10]. On-the-fly computation becomes especially important in online implementations via, *e.g.*, the *online EM* or *recursive maximum likelihood* approaches [6, 21].

As closed-form solutions to this smoothing problem can be obtained only for linear Gaussian models or models with finite state spaces $(\mathsf{X}_n)_{n \in \mathbb{N}}$, loads of papers have been written over the years with the aim of developing SMC-based approximative solutions. Most of these works assume that each density ℓ_n (or, in the HMM case, the transition density of $\mathbf{Q}_n$ and the likelihood g_n) is available in

a closed form; however, this is not the case for a large number of interesting models, including most state-space HMMs governed by stochastic differential equations. Still, there are a few exceptions in the literature. In [17] (see also [18]), the authors showed that asymptotically consistent online state estimation in partially observed diffusion processes can be achieved by means of a *random-weight particle filter*, in which unavailable importance weights are replaced by unbiased estimates (produced using so-called *generalized Poisson estimators* [4]). This approach is closely related to *pseudo-marginal methods* [1], since the unbiasedness allows the true, intractable target to be embedded into an extended distribution having the target as a marginal; as a consequence, the consistency of the algorithm follows straightforwardly from standard SMC convergence results. A similar pseudo-marginal SMC approach was developed in [25] for random effects models with non-analytic likelihood. In [27], this technology was cast into the framework of *fixed-lag particle smoothing* of additive state functionals, where the well-known particle-path degeneracy of naive particle smoothers is avoided at the price of a lag-induced bias. Recently, [38] designed a random-weight version of the forward-only particle smoother proposed in [10], whose computational complexity is quadratic in the number N of particles, yielding a strongly consistent—though computationally demanding—algorithm. Moreover, [19] extended the random-weight particle filtering approach to the *particle-based, rapid incremental smoother* (PaRIS), proposed in [28] as a means for additive smoothing in HMMs, yielding an algorithm with just linear complexity. The complexity of the latter algorithm is appealing; however, the schedule was not furnished with any theoretical results concerning the asymptotic properties and long-term stability of the estimator or the effect of the weight randomisation on the accuracy. In addition, the algorithm is restricted to partially observed diffusions and unbiased weight estimation, calling for strong assumptions on the unobserved process.

In the present paper we further develop the approach in [19] and extend the PaRIS to online additive smoothing in general models in the form (1.1) and the scenario where the transition densities $(\ell_n)_{n \in \mathbb{N}}$ are intractable but can be estimated by means of simulation. These estimates may be unbiased or biased. In its original form, the PaRIS avoids particle-path degeneracy by alternating two sampling operations, one that propagates a sample of forward-filtering particles and another that resamples a set of backward-smoothing statistics, and the proposed method replaces the sampling distributions associated with these operations by suitable pseudo-marginals. This leads to an $\mathcal{O}(N)$ algorithm that can be applied to a wide range of smoothing problems, including additive smoothing in partially observed diffusion processes and additive *approximate Bayesian computation smoothing* [24]. As illustrated by our examples, it covers the random-weight algorithms proposed in [17] and [19] as special cases and provides, as another special case, an extension of the original PaRIS proposed in [28] to general path models (1.1) and auxiliary particle filters. In addition, the proposed method is furnished with a rigorous theoretical analysis, the results of which can be summarised as follows.

- We establish exponential concentration and asymptotic normality of the estimators produced by the algorithm. These results extend analogous results established in [28] for the original PaRIS (operating on fully dominated HMMs using the bootstrap particle filter), and the additional randomness of the pseudo-marginals can be shown to effect the asymptotic variance through an additional positive term. The fact that our smoothing algorithm, as explained above, involves two separate levels of pseudo-marginalisation makes this extension highly non-trivial.
- Under strong mixing assumptions we establish the long-term stochastic stability of our algorithm by showing that its asymptotic variance grows at most linearly in n. As explained in [28, Section 1], this is optimal for a path-space Monte Carlo estimator. As a by-product of this analysis, we obtain a time-uniform bound on the asymptotic variance of the random-weight particle filter.
- As mentioned above, we do not require the estimators of $(\mathbf{L}_n)_{n \in \mathbb{N}}$ to be unbiased. The bias is assumed to be regulated by some precision parameter ε (see (**H4**)), and under additional strong mixing assumptions we establish an $\mathcal{O}(n\varepsilon)$ bound on the asymptotic bias of the final estimator. In addition, we obtain, as a by-product, an $\mathcal{O}(\varepsilon)$ bound for the random-weight particle filter. These results are the first of its kind.

The paper is structured as follows. In Section 2 we cast, under the temporary assumption that each ℓ_n is tractable, the PaRIS into the general model (1.1) and auxiliary particle filters and define carefully the two forward and backward sampling operations constituting the algorithm. Since this extension is of independent interest, we provide the details. In Section 3 we show how pseudo-marginal forward and backward sampling allow the temporary tractability assumption to be abandoned. Section 4 presents all theoretical results and although an extensive numerical study of the proposed scheme is beyond the scope of our paper, we present a minor numerical illustration of the $\mathcal{O}(n\varepsilon)$ bias bound in Section 5. Sections A–E contain all proofs.

2. Preliminaries

We first introduce some general notation. For any $(m, n) \in \mathbb{Z}$ such that $n \leq m$, we let $[\![m, n]\!]$ denote the set $\{m, \ldots, n\}$. For arbitrary elements a_ℓ, $\ell \in [\![m, n]\!]$, we denote vectors by $a_{m:n} = (a_m, \ldots, a_n)$. The sets of measures, probability measures, and real-valued bounded measurable functions on some given state-space $(\mathsf{X}, \mathcal{X})$ are denoted by $\mathsf{M}(\mathcal{X})$, $\mathsf{M}_1(\mathcal{X})$, and $\mathsf{F}(\mathcal{X})$, respectively. For any measure μ and measurable function h we let $\mu h \coloneqq \int h(x) \, \mu(dx)$ denote the Lebesgue integral of h with respect to μ whenever this is well defined. We will write $\mu^2 f = (\mu f)^2$ (whereas $\mu f^2 = \mu(f^2)$). For any finite set S, $\mathcal{P}(S)$ denotes the power set of S. The following kernel notation will be used repeatedly in the paper. Let $(\mathsf{X}, \mathcal{X})$ and $(\mathsf{Y}, \mathcal{Y})$ be general state spaces and $\mathbf{K} : \mathsf{X} \times \mathcal{Y} \to \mathbb{R}_{\geq 0}$ some transition kernel. Then $\mathbf{K}$ induces two operators, one acting on measurable functions and the other on measures. More precisely, for any $h \in \mathsf{F}(\mathcal{X} \otimes \mathcal{Y})$ and

$\mu \in \mathsf{M}(\mathcal{X})$, let

$$\mathbf{K}h : \mathsf{X} \ni x \mapsto \int h(x,y)\,\mathbf{K}(x,dy), \quad \mu\mathbf{K} : \mathcal{Y} \ni A \mapsto \int \mu(dx)\,\mathbf{K}(x,A)$$

Moreover, let $(\mathsf{Z},\mathcal{Z})$ be a third state space and $\mathbf{K}' : \mathsf{Y} \times \mathcal{Z} \to \mathbb{R}_{\geq 0}$ another kernel; then the product of $\mathbf{K}$ and $\mathbf{K}'$ is the kernel defined by

$$\mathbf{K}\mathbf{K}' : (x,A) \ni \mathsf{X} \times \mathcal{Z} \mapsto \int \mathbf{K}(x,dy)\,\mathbf{K}'(y,A).$$

2.1. Model and aim

With notations as in Section 1, define, for each $n \in \mathbb{N}$ and $m \in [\![0,n]\!]$, the kernel

$$\mathbf{L}_{m,n}(x'_{0:m},dx_{0:n+1}) \coloneqq \delta_{x'_{0:m}}(dx_{0:m}) \prod_{\ell=m}^{n} \mathbf{L}_\ell(x_\ell,dx_{\ell+1}) \tag{2.1}$$

on $\mathsf{X}^n \times \mathcal{X}^{n+1}$. In addition, let $\mathbf{L}_{n,n-1} = \mathrm{id}$. Note that $\mathbf{L}_{n,n}$ is different from $\mathbf{L}_n$ in the sense that the former is defined on $\mathsf{X}^n \times \mathcal{X}^{n+1}$ whereas the latter is defined on $\mathsf{X}_n \times \mathcal{X}_{n+1}$. We will always assume that for all $n \in \mathbb{N}$, $\chi\mathbf{L}_{0,n-1}\mathbb{1}_{\mathsf{X}^n} = \chi\mathbf{L}_0 \cdots \mathbf{L}_{n-1}\mathbb{1}_{\mathsf{X}_n} > 0$. Since each mapping $\mathbf{L}_{m,n}\mathbb{1}_{\mathsf{X}^n}$ depends only on the last coordinate x_m, a version of this mapping with domain X_m is well defined; we will denote the latter by the same symbol and write $\mathbf{L}_{m,n}\mathbb{1}_{\mathsf{X}^n}(x_m)$, $x_m \in \mathsf{X}_m$, when needed. Using the previous notations, the path measures (1.1) can be expressed as

$$\phi_{0:n}(dx_{0:n}) = \frac{\chi\mathbf{L}_{0,n-1}(dx_{0:n})}{\chi\mathbf{L}_{0,n-1}\mathbb{1}_{\mathsf{X}^n}}, \quad n \in \mathbb{N}. \tag{2.2}$$

For each $n \in \mathbb{N}$, let $\phi_n : A \ni \mathcal{X}_n \mapsto \phi_{0:n}(\mathsf{X}^{n-1} \times A)$ denote the marginal of $\phi_{0:n}$ with respect to the last component. Note that the path- and marginal-measure flows can be expressed recursively as

$$\phi_{0:n+1}(dx_{0:n+1}) = \frac{\phi_{0:n}\mathbf{L}_{n,n}(dx_{0:n+1})}{\phi_{0:n}\mathbf{L}_{n,n}\mathbb{1}_{\mathsf{X}^{n+1}}} = \frac{\phi_{0:n}\mathbf{L}_{n,n}(dx_{0:n+1})}{\phi_n\mathbf{L}_n\mathbb{1}_{\mathsf{X}_{n+1}}}, \quad n \in \mathbb{N}, \tag{2.3}$$

and

$$\phi_{n+1}(dx_{n+1}) = \frac{\phi_n\mathbf{L}_n(dx_{n+1})}{\phi_n\mathbf{L}_n\mathbb{1}_{\mathsf{X}_{n+1}}}, \quad n \in \mathbb{N}, \tag{2.4}$$

respectively. Given some sequence $(\tilde{h}_n)_{n\in\mathbb{N}}$ of functions $\tilde{h}_n : \mathsf{X}_n \times \mathsf{X}_{n+1} \to \mathbb{R}$, our aim is, as declared in Section 1, the online approximation of $(\phi_{0:n}h_n)_{n\in\mathbb{N}}$, where h_n defined by (1.2).

Remark 2.1. Note that our framework is equivalent with the Feynman-Kac models considered in [9, Section 1.3], where it is assumed that each kernel $\mathbf{L}_n$ can be decomposed into a Markov transition kernel $\mathbf{M}_n$ on $\mathsf{X}_n \times \mathcal{X}_{n+1}$ and a potential function $g_n \in \mathsf{F}(\mathcal{X}_n)$ according to $\mathbf{L}_n(x_n,dx_{n+1}) = g_n(x_n)\,\mathbf{M}(x_n,dx_{n+1})$.

Indeed, as soon as $\mathbf{L}_n$ is bounded, such a decomposition is always possible by letting $\mathbf{M}(x_n, dx_{n+1}) \coloneqq \mathbf{L}_n(x_n, dx_{n+1})/\mathbf{L}_n(x_n, \mathsf{X}_{n+1})$ and $g_n(x_n) \coloneqq \mathbf{L}_n(x_n, \mathsf{X}_{n+1})$. However, in our case this potential function is, on the contrary to what is assumed in [9], generally intractable, since $\mathbf{L}_n(x_n, \mathsf{X}_{n+1})$ is typically unknown for the models that we will consider. Moreover, as noted in [9, Section 1.3], the path model $(\phi_{0:n})_{n \in \mathbb{N}}$ and the marginal model $(\phi_n)_{n \in \mathbb{N}}$ have the same mathematical structure in the sense that the path model can be formulated as a marginal model evolving on the spaces $(\mathsf{X}'_n, \mathcal{X}'_n)_{n \in \mathbb{N}}$, where $\mathsf{X}'_n \coloneqq \mathsf{X}^n$ and $\mathcal{X}'_n \coloneqq \mathcal{X}^n$, according to the initial distribution $\chi \coloneqq \chi'$ and the transition kernels $(\mathbf{L}'_n)_{n \in \mathbb{N}}$, where $\mathbf{L}'_n \coloneqq \mathbf{L}_{n,n}$. Still, the kernels $(\mathbf{L}'_n)_{n \in \mathbb{N}}$ involve transitions according Dirac measures, which makes the model formed by χ' and $(\mathbf{L}'_n)_{n \in \mathbb{N}}$ ill-suited for naive particle approximation; see Section 2.2.1 for further discussion.

Example 1 (state-space models). Let $(\mathsf{X}, \mathcal{X})$ and $(\mathsf{Y}, \mathcal{Y})$ be general state spaces and $(\mathbf{Q}_n)_{n \in \mathbb{N}}$ and $(\mathbf{G}_n)_{n \in \mathbb{N}}$ sequences of Markov kernels on $\mathsf{X} \times \mathcal{X}$ and $\mathsf{X}^2 \times \mathcal{Y}$, respectively. In addition, let χ be some probability measure on $\mathcal{X}$. Consider a fully dominated model where all $\mathbf{G}_n$ and $\mathbf{Q}_n$ have transition densities g_n and q_n with respect to some reference measures ν and μ on $\mathcal{Y}$ and $\mathcal{X}$, respectively. Let $\{X_0, (X_n, Y_n) : n \in \mathbb{N}_{>0}\}$ be the canonical Markov chain induced by the initial distribution χ and the Markov kernel $\mathbf{Q}_n(x_n, dx_{n+1})\, \mathbf{G}_n(x_n, x_{n+1}, dy_{n+1})$ (which has no dependence on the y_n variable; the same dynamics hence applies to the first transition $X_0 \rightsquigarrow (X_1, Y_1)$) and denote by $\mathbb{P}_\chi$ its law with corresponding expectation $\mathbb{E}_\chi$. In this model, we assume that the *state process* $(X_n)_{n \in \mathbb{N}}$ is latent and only partially observed thought the *observation process* $(Y_n)_{n \in \mathbb{N}_{>0}}$. It can be shown that (i) the state process is itself a Markov chain with initial distribution χ and transition kernels $(\mathbf{Q}_n)_{n \in \mathbb{N}}$ and (ii) conditionally to the state process, the observations are independent and such that the marginal distribution of Y_n is given by $\mathbf{G}_{n-1}(X_{n-1}, X_n, \cdot)$ for all n. In the case where $\mathbf{G}_{n-1}$ does not depend on x_{n-1}, the model is a fully adapted general state-space HMM; see [7, Section 2.2]. In this setting, the joint-smoothing distribution at time $n \in \mathbb{N}$ is, for a given record $y_{1:n} \in \mathsf{Y}^n$ of observations, defined as the probability measure

$$\phi_{0:n}\langle y_{1:n}\rangle(dx_{0:n}) \coloneqq L_n^{-1}(y_{1:n})\chi(dx_0) \prod_{m=0}^{n-1} \mathbf{Q}_m(x_m, dx_{m+1})\, g_m(x_m, x_{m+1}, y_{m+1}),$$
(2.5)

on $\mathcal{X}^n$, where

$$L_n(y_{1:n}) \coloneqq \int \cdots \int \chi(dx_0) \prod_{m=0}^{n-1} \mathbf{Q}_m(x_m, dx_{m+1})\, g_m(x_m, x_{m+1}; y_{m+1}) \quad (2.6)$$

is the observed data likelihood. Along the lines of [7, Proposition 3.1.4] one may show that $\phi_{0:n}\langle Y_{1:n}\rangle$ is, $\mathbb{P}_\chi$-a.s., the conditional distribution of $X_{0:n}$ given $Y_{1:n}$. The marginal $\phi_n\langle y_{1:n}\rangle$ of the joint smoothing distribution with respect to its last component x_n is referred to as the *filtering distribution at time n*. Consequently, by defining kernel densities $\ell_n(x_n, x_{n+1}) \coloneqq g_n(x_n, x_{n+1}, y_{n+1})q_n(x_n, x_{n+1})$ for all $n \in \mathbb{N}$ (while keeping dependence on observations implicit) and letting $\mathbf{L}_n$

be the induced transition kernels, the joint-smoothing distributions may be expressed in the form (2.2).

The measures (2.2) are generally intractable in two ways. First, in many applications the transition densities $(\ell_n)_{n\in\mathbb{N}}$ cannot be evaluated pointwise. Returning to Example 1 and the context of smoothing in state-space models, this is typically the case when the dynamics of the latent process is governed by a stochastic differential equation (see Example 2 below). Second, even in the case where these transition densities are evaluable, the normalising constant in (2.2) is generally intractable. Thus, in order to solve the smoothing problem in full generality, one needs to be able to handle this double intractability, which is the goal of the algorithm that we will develop next. We will proceed in two steps. In the next section, Section 2.2, we will solve the additive smoothing problem under the temporary assumption that the densities $(\ell_n)_{n\in\mathbb{N}}$ are tractable; the resulting extension of the PaRIS proposed in [28] to general models in the form (1.1) and auxiliary particle filters is of independent interest. Then, in Section 3, we abandon the temporary assumption of tractability and assume that the user has only possibly biased proxies of these densities at hand.

2.2. The PaRIS

2.2.1. Auxiliary particle filters

Assume for a moment that each transition density ℓ_n is available in a closed form. Then standard SMC methods (see [8] for a recent introduction) can be used to approximate the distribution flows $(\phi_{0:n})_{n\in\mathbb{N}}$ and $(\phi_n)_{n\in\mathbb{N}}$ using Monte Carlo samples generated recursively by means of sequential importance sampling and resampling operations. In order to set notations, let us recall the most general class of such algorithms, the so-called auxiliary particle filters [31]. In the light of Remark 2.1 it is enough to consider particle approximation of the marginals $(\phi_n)_{n\in\mathbb{N}}$. We proceed recursively and assume that we, at time $n \in \mathbb{N}$, have at hand a sample $(\xi_n^i, \omega_n^i)_{i=1}^N$ of X_n-valued *particles* (the ξ_n^i) and associated nonnegative importance weights (the ω_n^i) such that the self-normalised estimator $\phi_n^N h \coloneqq \Omega_n^{-1} \sum_{i=1}^N \omega_n^i h(\xi_n^i)$, with $\Omega_n \coloneqq \sum_{i=1}^N \omega_n^i$, approximates $\phi_n h$ for every ϕ_n-integrable function h. Then plugging ϕ_n^N into the recursion (2.4) yields the approximation $\sum_{i=1}^N \pi_n(i, dx)$ of ϕ_{n+1}, where

$$\pi_n(i, dx) \propto \omega_n^i \mathbf{L}_n(\xi_n^i, dx) \tag{2.7}$$

is a mixture distribution on $\mathcal{P}(\llbracket 1, N \rrbracket) \otimes \mathcal{X}_{n+1}$. In order to form new particles approximating ϕ_{n+1}, we may draw, using importance sampling, pairs $(I_{n+1}^i, \xi_{n+1}^i)_{i=1}^N$ of indices and particles from π_n and discard the former. For this purpose, we introduce some instrumental mixture distribution

$$\rho_n(i, dx) \propto \omega_n^i \vartheta_n(\xi_n^i) \mathbf{P}_n(\xi_n^i, dx) \tag{2.8}$$

on the same space, where ϑ_n is a real-valued positive *adjustment-weight function* on X_n and $\mathbf{P}_n$ is a proposal kernel on $\mathsf{X}_n \times \mathcal{X}_{n+1}$ such that $\mathbf{L}_n(x, \cdot) \ll \mathbf{P}_n(x, \cdot)$

for all $x \in \mathsf{X}_n$. We will always assume that $\mathbf{P}_n$ has a transition density p_n with respect to μ_{n+1}. A draw (I_{n+1}^i, ξ_{n+1}^i) from ρ_n is easily generated by first drawing I_{n+1}^i from the categorical distribution induced by the adjusted importance weights and then drawing ξ_{n+1}^i by moving randomly the selected ancestor $\xi_n^{I_{n+1}^i}$ according to the proposal kernel. These steps are often referred to as *selection* and *mutation*, respectively. Finally, each draw ξ_{n+1}^i is assigned the updated importance weight proportional to $d\pi_n/d\rho_n(I_{n+1}^i, \xi_{n+1}^i)$, and the estimator $\phi_{n+1}^N h = \Omega_{n+1}^{-1} \sum_{i=1}^N \omega_{n+1}^i h(\xi_{n+1}^i)$ approximates $\phi_{n+1}h$ for every ϕ_{n+1}-integrable h. The full update, which we will refer to as *forward sampling* and express in a short form as,

$$(\xi_{n+1}^i, \omega_{n+1}^i)_{i=1}^N \sim \mathsf{FS}((\xi_n^i, \omega_n^i)_{i=1}^N), \tag{2.9}$$

is summarised in Algorithm 1.[2]

Data: $(\xi_n^i, \omega_n^i)_{i=1}^N$
Result: $(\xi_{n+1}^i, \omega_{n+1}^i)_{i=1}^N$

1 **for** $i = 1 \to N$ **do**
2   draw $I_{n+1}^i \sim \mathsf{cat}(\{\vartheta_n(\xi_n^\ell)\omega_n^\ell\}_{\ell=1}^N)$;
3   draw $\xi_{n+1}^i \sim \mathbf{P}_n(\xi_n^{I_{n+1}^i}, \cdot)$;
4   set $\omega_{n+1}^i \leftarrow \dfrac{\ell_n(\xi_n^{I_{n+1}^i}, \xi_{n+1}^i)}{\vartheta_n(\xi_n^{I_{n+1}^i})p_n(\xi_n^{I_{n+1}^i}, \xi_{n+1}^i)}$;

Algorithm 1: Forward sampling, FS

With this terminology, the auxiliary particle filter consists of iterated forward sampling operations, and we will assume that the process is initialised by sampling independent particles $(\xi_0^i)_{i=1}^N$ from some proposal ν on $(\mathsf{X}_0, \mathcal{X}_0)$ such that $\chi \ll \nu$ and letting $\omega_0^i := d\chi/d\nu(\xi_0^i)$ for all i.

Note that we may, in the light of Remark 2.1, obtain particle approximations $(\xi_{0:n}^i, \omega_n^i)_{i=1}^N$, $n \in \mathbb{N}$, of the smoothing distribution flow $(\phi_{0:n})_{n \in \mathbb{N}}$ by applying the previous sampling scheme to the model formed by χ' and $(\mathbf{L}_n')_{n \in \mathbb{N}}$. From an algorithmic point of view, it is easy to see that the only change needed is to insert, just after Line 3, the command $\xi_{0:n+1}^i \leftarrow (\xi_{0:n}^{I_{n+1}^i}, \xi_{n+1}^i)$ storing the particle paths (in particular, the weight-updating step on Line 4 remains the same). Still, it is well known that repeated selection operations lead to coalescing paths $(\xi_{0:n}^i)_{i=1}^N$, and as a consequence the variance of this naive smoothing estimator increases rapidly with n; indeed, in the case of additive state functionals, the growth in variance is typically quadratic in n (see [28] for a discussion), which is unreasonable from a computational point of view. We will thus rely on more advanced, stochastically stable smoothing technology avoiding the particle-path degeneracy problem by taking advantage of the time-uniform convergence of the marginal samples $(\xi_n^i)_{i=1}^N$. This will be discussed in the next section.

---

[2]Mathematically, the forward sampling operation defines a Markov transition kernel, which motivates the use of the symbol $\sim$ in (2.9).

Finally, we note that the re-weighting operation on Line 4 in Algorithm 1 requires the transition density ℓ_n to be tractable, which is not the case in general. We will return to the general case in Section 3.

2.2.2. Backward sampling

The following quantities will play a key role in the following. For each $m \in \mathbb{N}$, define the *backward Markov kernel*

$$\mathbf{B}_m(x_{m+1}, dx_m) := \frac{\phi_m(dx_m)\, \ell_m(x_m, x_{m+1})}{\phi_m[\ell_m(\cdot, x_{m+1})]} \tag{2.10}$$

on $\mathsf{X}_{m+1} \times \mathcal{X}_m$. In addition, for each $n \in \mathbb{N}_{>0}$, let the Markov kernel

$$\mathbf{T}_n(x_n, dx_{0:n-1}) := \prod_{m=0}^{n-1} \mathbf{B}_m(x_{m+1}, dx_m) \tag{2.11}$$

on $\mathsf{X}_n \times \mathcal{X}^{n-1}$ denote the joint law of the backward Markov chain induced by the kernels (2.10) when initialised at $x_n \in \mathsf{X}_n$. An important class of sequential Monte Carlo joint-smoothing methods [15, 20] is based on the following result.

Lemma 2.2.

(i) For all $n \in \mathbb{N}$ and $h \in \mathsf{F}(\mathcal{X}_n \otimes \mathcal{X}_{n+1})$,

$$\iint \phi_n(dx_n)\, \mathbf{L}_n(x_n, dx_{n+1})\, h(x_n, x_{n+1})$$
$$= \iint \phi_n \mathbf{L}_n(dx_{n+1})\, \mathbf{B}_n(x_{n+1}, dx_n)\, h(x_n, x_{n+1}). \tag{2.12}$$

(ii) For all $n \in \mathbb{N}_{>0}$ and $h \in \mathsf{F}(\mathcal{X}^n)$, $\phi_{0:n} h = \phi_n \mathbf{T}_n h$.

In the case of state-space models, the identity (ii) above is a well-known result typically referred to as the *backward decomposition* of the joint-smoothing distribution; still, as far as known to the authors, it has never been established in the general setting considered in the present paper, and a proof of Lemma 2.2 is hence given in Section A for completeness. Importantly, as noted in [6], the functions $(\mathbf{T}_n h_n)_{n \in \mathbb{N}_{>0}}$ can be expressed recursively through

$$\mathbf{T}_{n+1} h_{n+1}(x_{n+1}) = \int (\mathbf{T}_n h_n(x_n) + \tilde{h}_n(x_n, x_{n+1}))\, \mathbf{B}_n(x_{n+1}, dx_n), \quad n \in \mathbb{N}, \tag{2.13}$$

with, by convention, $\mathbf{T}_0 h_0 \equiv 0$. Here the backward kernel $\mathbf{B}_n$ depends on the marginal ϕ_n; thus, the recursion is driven by the marginal flow $(\phi_n)_{n \in \mathbb{N}}$, which may again be expressed recursively through (2.4). However, as these marginals are, as already mentioned, generally intractable, exact computations need typically to be replaced by approximations. The authors of [10] propose to approximate the values of each statistic $\mathbf{T}_n h_n$ at a random, discrete support formed by

particles. More precisely, assume again that the transition density ℓ_n is tractable and, by induction, that we at time step n have at hand a given particle sample $(\xi_n^i, \omega_n^i)_{i=1}^N$ and a set of statistics $(\tau_n^i)_{i=1}^N$ such that τ_n^i is an approximation of $\mathbf{T}_n h_n(\xi_n^i)$. Then, in order to propagate the statistics $(\tau_n^i)_{i=1}^N$ forward, one updates, in a first substep, the particle sample $(\xi_n^i, \omega_n^i)_{i=1}^N$ recursively by forward sampling (Algorithm 1). After forward sampling, one replaces, in the definition (2.10) of $\mathbf{B}_n$, ϕ_n by the corresponding particle approximation, yielding the updates

$$\tau_{n+1}^i = \sum_{j=1}^N \mathbf{\Pi}_n(i,j)(\tau_n^j + \tilde{h}_n(\xi_n^j, \xi_{n+1}^i)), \quad i \in [\![1, N]\!], \tag{2.14}$$

where we have defined the transition kernel

$$\mathbf{\Pi}_n(i,j) := \frac{\omega_n^j \ell_n(\xi_n^j, \xi_{n+1}^i)}{\sum_{j'=1}^N \omega_n^{j'} \ell_n(\xi_n^{j'}, \xi_{n+1}^i)} \tag{2.15}$$

on $[\![1, N]\!] \times \mathcal{P}([\![1, N]\!])$. Since computing each τ_n^i according to (2.14) has a linear computational complexity in the number N of particles, the overall complexity this approach is *quadratic* in N. In order to deal with this significant computational burden, the authors of [28] suggest replacing summation by additional Monte Carlo simulation. More precisely, by sampling, for each i, $M \in \mathbb{N}_{>0}$ independent indices $(J^{(i,j)})_{j=1}^M$ from $\mathbf{\Pi}_n(i, \cdot)$ and replacing (2.14) by

$$\tau_{n+1}^i = \frac{1}{M} \sum_{j=1}^M \left(\tau_n^{J_{n+1}^{(i,j)}} + \tilde{h}_n(\xi_n^{J_{n+1}^{(i,j)}}, \xi_{n+1}^i) \right), \quad i \in [\![1, N]\!], \tag{2.16}$$

the computational complexity can, as we shall soon see, be reduced significantly. At each iteration, the self-normalised estimator $\Omega_n^{-1} \sum_{i=1}^N \omega_n^i \tau_n^i$ serves as an estimator of the quantity $\phi_n \mathbf{T}_n h_n = \phi_{0:n} h_n$ of interest. This second operation, which we will refer to as *backward sampling*,

$$(\tau_{n+1}^i)_{i=1}^N \sim \mathsf{BS}((\xi_n^i, \tau_n^i, \omega_n^i)_{i=1}^N, (\xi_{n+1}^i)_{i=1}^N),$$

is summarised in Algorithm 2.

Data: $(\xi_n^i, \tau_n^i, \omega_n^i)_{i=1}^N$, $(\xi_{n+1}^i)_{i=1}^N$
Result: $(\tau_{n+1}^i)_{i=1}^N$
1 **for** $i = 1 \to N$ **do**
2   **for** $j = 1 \to M$ **do**
3   draw $J_{n+1}^{(i,j)} \sim \mathbf{\Pi}_n(i, \cdot)$;
4   set $\tau_{n+1}^i \leftarrow \frac{1}{M} \sum_{j=1}^M \left(\tau_n^{J_{n+1}^{(i,j)}} + \tilde{h}_n(\xi_n^{J_{n+1}^{(i,j)}}, \xi_{n+1}^i) \right)$;

Algorithm 2: Backward sampling, BS

Let us examine closer the sampling step on Line 3 in Algorithm 2. In order to keep the algorithmic complexity at a reasonable level, the computation of the

normalising constant of $\mathbf{\Pi}_n(i, \cdot)$, which consists of N terms, should be avoided; otherwise, the overall complexity remains quadratic in N. This is possible using, *e.g.*,

– *rejection sampling*. This approach relies on the mild assumption that there exists some measurable function c on X_{n+1} such that $\ell_n(x_n, x_{n+1}) \leq c(x_{n+1})$ for all $x_{n:n+1} \in \mathsf{X}_n \times \mathsf{X}_{n+1}$. Then, following [13], $\mathbf{\Pi}_n(i, \cdot)$ can be sampled from by generating a candidate J^* from $\mathsf{cat}(\{\omega_n^\ell\}_{\ell=1}^N)$ and accepting the same with probability

$$\alpha^{\mathsf{R}} := \frac{\ell_n(\xi_n^{J^*}, \xi_{n+1}^i)}{c(\xi_{n+1}^i)}. \tag{2.17}$$

The procedure is repeated until acceptance, conditionally to which J^* is distributed according to $\mathbf{\Pi}_n(i, \cdot)$. Since the $\mathsf{cat}(\{\omega_n^\ell\}_{\ell=1}^N)$ distribution is independent of i, this circumvents the need to compute a normalising sum for every i. The approach may reduce significantly the computational complexity; indeed, as shown in [13, Proposition 2], the expected overall complexity of the algorithm is *linear* in N under certain assumptions.

– *MCMC methods*. Another possibility is to generate the variables $(J_{n+1}^{(i,j)})_{j=1}^M$ using the Metropolis-Hastings algorithm. For this purpose, let ρ be some proposal transition density on $[\![1, N]\!]^2$. Then proceeding recursively, given $J_{n+1}^{(i,j)} = J$, a candidate J^* for $J_{n+1}^{(i,j+1)}$ is sampled from the density $\rho(J, \cdot)$ and accepted with probability

$$\alpha^{\mathsf{MH}} := 1 \wedge \frac{\omega_n^{J^*} \ell_n(\xi_n^{J^*}, \xi_{n+1}^i) \rho(J^*, J)}{\omega_n^J \ell_n(\xi_n^J, \xi_{n+1}^i) \rho(J, J^*)}.$$

If rejection, then $J_{n+1}^{(i,j+1)} = J$. It is close at hand is to let ρ take the form of an independent proposal given by the $\mathsf{cat}(\{\omega_n^i\}_{i=1}^N)$ distribution; in that case α^{MH} simplifies to

$$\alpha^{\mathsf{MH}} = 1 \wedge \frac{\ell_n(\xi_n^{J^*}, \xi_{n+1}^i)}{\ell_n(\xi_n^J, \xi_{n+1}^i)}. \tag{2.18}$$

With this approach, the variables $(J_{n+1}^{(i,j)})_{j=1}^M$ are conditionally dependent; this can be counteracted by including only an m-skeleton of this sequence in the update (2.16). An important advantage of this approach over rejection sampling is that it does not require ℓ_n to be dominated.

Finally, combining the forward and backward sampling operations in accordance with Algorithm 3 yields a generalisation of the PaRIS proposed in [28] to a general framework comprising Feynman-Kac models and auxiliary particle filters.

Algorithm 3 is initialised by drawing $(\xi_0^i)_{i=1}^N \sim \nu^{\otimes N}$ and letting $\omega_0^i = d\chi/d\nu(\xi_0^i)$ and $\tau_n^i = 0$.

In this scheme, the sample size M of the backward sampling operation is an algorithmic parameter that has to be set by the user. As shown in Section 4, the

Data: $(\xi_n^i, \tau_n^i, \omega_n^i)_{i=1}^N$
Result: $(\xi_{n+1}^i, \tau_{n+1}^i, \omega_{n+1}^i)_{i=1}^N$
1 run $(\xi_{n+1}^i, \omega_{n+1}^i)_{i=1}^N \sim \mathsf{FS}((\xi_n^i, \omega_n^i)_{i=1}^N)$;
2 run $(\tau_{n+1}^i)_{i=1}^N \sim \mathsf{BS}((\xi_n^i, \tau_n^i, \omega_n^i)_{i=1}^N, (\xi_{n+1}^i)_{i=1}^N)$;

Algorithm 3: Full PaRIS update.

produced estimators are, for all $n \in \mathbb{N}$, consistent and asymptotically normal for any fixed M larger than or equal to one. In addition, for any fixed $M \geq 2$ the algorithm is stochastically stable with an $\mathcal{O}(n)$ variance, which is optimal; see [28, Section 1] for a discussion. These results form a nontrivial extension of similar results obtained by [28] in the simpler setting of state-space models and bootstrap particle filters.

Finally, we remind the reader that we have here considered the idealised situation where the unnormalised transition densities $(\ell_n)_{n \in \mathbb{N}}$ can be evaluated pointwise, which will generally not be the case for the applications we will consider. Thus, in the next section we will approach the more general case where these transition densities are intractable but may be estimated, and we will show how consistent, asymptotically normal, and stochastically stable estimators can be produced also in such a scenario by pseudo-marginalising separately the forward and backward sampling operations.

3. Pseudo-marginal PaRIS algorithms

3.1. Pseudo marginalisation in Monte Carlo methods

Pseudo-marginalisation was originally proposed in [3] in the framework of MCMC methods, and in [1] the method was developed further and provided with a solid theoretical basis. In the following we recapitulate briefly the main idea behind this approach. Consider the problem of sampling from some target distribution π defined on some measurable space $(\mathsf{X}, \mathcal{X})$ and having a density with respect to some reference measure μ. This density is assumed to be proportional to some *intractable* nonnegative measurable function ℓ on X, *i.e.*, $\pi(dx) = \lambda(dx)/\lambda\mathbb{1}_\mathsf{X}$, where $\lambda(dx) \coloneqq \ell(x)\,\mu(dx)$ is finite. While the target density is intractable we assume that there exist some additional state space $(\mathsf{Z}, \mathcal{Z})$, a Markov kernel $\mathbf{R}$ on $\mathsf{X} \times \mathcal{Z}$, and some nonnegative measurable function $\mathsf{X} \times \mathsf{Z} \ni (x, z) \mapsto \ell\langle z\rangle(x)$ known up to a constant of proportionality and such that for all $x \in \mathsf{X}$,

$$\int \ell\langle z\rangle(x)\,\mathbf{R}(x, dz) = \ell(x). \tag{3.1}$$

Thus, a pointwise estimate of $\ell(x)$ can be obtained by generating ζ from $\mathbf{R}(x, dz)$ and computing the statistic $\ell\langle\zeta\rangle(x)$, the *pseudo marginal*. In Monte Carlo methods, the measure to replace, when necessary, the true marginal ℓ by its pseudo marginal is referred to as *pseudo marginalisation*. Interestingly, even though

pseudo marginalisation is based on the plug-in principle, it preserves typically the consistency of an algorithm. In order to see the this, let $\bar{\mathsf{X}} \coloneqq \mathsf{X} \times \mathsf{Z}$ and $\bar{\mathcal{X}} \coloneqq \mathcal{X} \otimes \mathcal{Z}$; then one may define an extended target distribution $\bar{\pi}(d\bar{x}) \coloneqq \bar{\lambda}(d\bar{x})/\bar{\lambda}\mathbb{1}_{\bar{\mathsf{X}}} = \bar{\lambda}(d\bar{x})/\lambda\mathbb{1}_{\mathsf{X}}$ on $(\bar{\mathsf{X}}, \bar{\mathcal{X}})$, where

$$\bar{\lambda}(d\bar{x}) \coloneqq \ell\langle z\rangle(x)\,\mathbf{R}(x, dz)\,\mu(dx) \tag{3.2}$$

(with $\bar{x} = (x, z)$). By (3.1), π is the marginal of $\bar{\pi}$ with respect to the x component. This means that we may produce a random sample $(\xi^i)_{i=1}^N$ in X targeting π by generating a sample $(\xi^i, \zeta^i)_{i=1}^N$ targeting $\bar{\pi}$ and simply discarding the Z-valued variables $(\zeta^i)_{i=1}^N$. Let ρ be a Markov transition density on X with respect to the reference measure μ. Then following [1], a Markov chain $(\xi_m, \zeta_m)_{m \in \mathbb{N}}$ targeting $\bar{\pi}$ can be produced on the basis of the Metropolis-Hastings algorithm as follows. Given a state (ξ_m, ζ_m), a candidate (ξ^*, ζ^*) for the next state is generated by drawing $\xi^* \sim \rho(x)\,\mu(dx)$ and $\zeta^* \sim \mathbf{R}(\xi^*, dz)$ and accepting the same with probability

$$\alpha \coloneqq 1 \wedge \frac{\ell\langle\zeta^*\rangle(\xi^*)\rho(\xi^*, \xi_m)}{\ell\langle\zeta_m\rangle(\xi_m)\rho(\xi_m, \xi^*)},$$

which is tractable. Note that α is indeed a pseudo-marginal version of the exact acceptance probability $1 \wedge \ell(\xi^*)\rho(\xi^*, \xi_m)/\ell(\xi_m)\rho(\xi_m, \xi^*)$ obtained if ℓ would be known. Note that the auxiliary variable ζ enters α only through the estimates $\ell\langle\zeta^*\rangle(\xi^*)$ and $\ell\langle\zeta_m\rangle(\xi_m)$, and since the latter has already been computed at the previous iteration there is no need of recomputing this quantity. In the *Monte-Carlo-within-Metropolis algorithm* (see again [1]) $\ell\langle z\rangle(x)$ is a pointwise importance sampling estimate of $\ell(x)$ based on a Monte Carlo sample ζ generated from $\mathbf{R}(x, \cdot)$. Alternatively, the extended distribution $\bar{\pi}$ can be sampled using rejection sampling or importance sampling, leading to consistent pseudo-marginal formulations of these algorithms as well.

In the present paper we will generalise the pseudo-marginal approach towards biased estimation by allowing the function

$$\ell^\varepsilon(x) \coloneqq \int \ell\langle z\rangle(x)\,\mathbf{R}(x, dz) \tag{3.3}$$

on X to be different from ℓ. Here $\varepsilon \geq 0$ is some accuracy parameter describing the distance between ℓ^ε and ℓ. Such biased estimates appear naturally when the law π is, *e.g.*, governed by a diffusion process and the density ℓ is approximated on the basis of different discretisation schemes; see the next section. In that case, ε plays the role of the discretisation step size. In (3.3), also the estimator $\ell\langle z\rangle(x)$ and the kernel $\mathbf{R}(x, dz)$ may depend on ε, even though this is suppressed in the notation. By introducing the possibly unnormalised measure $\lambda^\varepsilon(dx) \coloneqq \ell^\varepsilon(dx)\,\mu(dx)$ on $\mathcal{X}$ we may define the *skew* target probability measure

$$\pi^\varepsilon(dx) \coloneqq \frac{\lambda^\varepsilon(dx)}{\lambda^\varepsilon\mathbb{1}_{\mathsf{X}}}$$

on $\mathcal{X}$. In the biased case, generating a sample $(\xi^i, \zeta^i, \omega^i)_{i=1}^N$ targeting (3.2) will, by (3.3), provide a sample $(\xi^i, \omega^i)_{i=1}^N$ targeting π^ε as a by-product. Thus, of

outmost importance is to obtain control over the bias between π and π^ε, which is possible under the assumption that there exists a constant $c \geq 0$ such that for all $h \in \mathsf{F}(\mathcal{X})$ and ε,

$$|\lambda^\varepsilon h - \lambda h| \leq c\varepsilon\|h\|_\infty. \tag{3.4}$$

For instance, in the diffusion process case mentioned above, a condition of type (3.4) holds, as we will see in Section 4, typically true in the case where the density is approximated using the *Durham-Gallant estimator* [16]. Using that for all $h \in \mathsf{F}(\mathcal{X})$,

$$\pi^\varepsilon h - \pi h = \pi^\varepsilon h \left(1 - \frac{\lambda^\varepsilon \mathbb{1}_\mathsf{X}}{\lambda\mathbb{1}_\mathsf{X}}\right) + \frac{\lambda^\varepsilon h - \lambda h}{\lambda\mathbb{1}_\mathsf{X}},$$

we straightforwardly obtain the bound

$$|\pi^\varepsilon h - \pi h| \leq \varepsilon\frac{2c}{\lambda\mathbb{1}_\mathsf{X}}\|h\|_\infty$$

on the systematic error induced by the skew model. Note that the unbiased case (3.1) corresponds to letting $\varepsilon = 0$ in assumption (3.4). In the next section we will present a solution to the main problem addressed in Section 2.1 exploring a pseudo-marginalised version of the PaRIS discussed in Section 2.2.

3.2. Pseudo-marginal PaRIS

The algorithm that we will propose relies on the following assumption.

(**H1**) Let $(\mathsf{Z}_n, \mathcal{Z}_n)_{n\in\mathbb{N}_{>0}}$ be a sequence of general state spaces. For each $n \in \mathbb{N}$ there exist a Markov kernel $\mathbf{R}_n$ on $\mathsf{X}_n \times \mathsf{X}_{n+1} \times \mathcal{Z}_{n+1}$ and a positive measurable function $\ell_n\langle z\rangle(x_n, x_{n+1})$ on $\mathsf{X}_n \times \mathsf{X}_{n+1} \times \mathsf{Z}_{n+1}$ such that for every $x_{n:n+1} \in \mathsf{X}_n \times \mathsf{X}_{n+1}$, drawing $\zeta \sim \mathbf{R}_n(x_{n:n+1}, dz)$ and computing $\ell_n\langle\zeta\rangle(x_n, x_{n+1})$ yields an estimate of $\ell_n(x_n, x_{n+1})$.

Under (**H1**), we denote, for every $n \in \mathbb{N}$ and $x_{n:n+1} \in \mathsf{X}_n \times \mathsf{X}_{n+1}$, by

$$\ell_n^\varepsilon(x_n, x_{n+1}) := \int \mathbf{R}_n(x_{n:n+1}, dz)\,\ell_n\langle z\rangle(x_n, x_{n+1}) \tag{3.5}$$

the expectation of the statistic $\ell_n\langle\zeta\rangle(x_n, x_{n+1})$. Here ε is an accuracy parameter belonging to some parameter space $\mathcal{E} \subset \mathbb{R}$ and controlling the bias of the estimated model with respect to the true model; this will be discussed in depth in Section 4.3. For each $n \in \mathbb{N}$ we define the unnormalised kernel

$$\mathbf{L}_n^\varepsilon(x_n, dx_{n+1}) = \ell_n^\varepsilon(x_n, x_{n+1})\,\mu_{n+1}(dx_{n+1}) \tag{3.6}$$

on $\mathsf{X}_n \times \mathcal{X}_{n+1}$.

3.2.1. Pseudo-marginal forward sampling

In the case where each ℓ_n is intractable, so is the mixture distribution π_n defined in (2.7). Under (**H1**), we may, as in Section 3.1, aim at consistent pseudo-marginalisation of the forward-sampling operation by applying self-normalised importance sampling to the extended mixture

$$\bar{\pi}_n(i, dx, dz) \propto \omega_n^i \ell_n \langle z \rangle (\xi_n^i, x) \, \mu_{n+1}(dx) \, \mathbf{R}_n(\xi_n^i, x, dz)$$

on $\bar{\mathcal{X}}_{n+1} \coloneqq \mathcal{P}(\llbracket 1, N \rrbracket) \otimes \mathcal{X}_{n+1} \otimes \mathcal{Z}_{n+1}$ using the instrumental distribution

$$\bar{\rho}_n(i, dx, dz) \propto \omega_n^i \vartheta_n(\xi_n^i) \, \mathbf{P}_n(\xi_n^i, dx) \, \mathbf{R}_n(\xi_n^i, x, dz)$$

on the same space. Note that the marginal of $\bar{\pi}_n$ with respect to (i, x) is the distribution proportional to $\omega_n^i \mathbf{L}_n^\varepsilon(\xi_n^i, dx)$, whose distance to the target π_n of interest is controlled by the precision parameter ε. Here the adjustment multiplier ϑ_n and the proposal kernel $\mathbf{P}_n$ of the instrumental distribution are as in Section 2.2.1. Each draw from $\bar{\rho}_n$ is assigned an importance weight given by the (tractable) Radon–Nikodym derivative of $\bar{\pi}_n$ to $\bar{\rho}_n$. It is easily seen that this sampling operation, which is detailed in Algorithm 4, corresponds to replacing the intractable transition density ℓ_n on Line 4 in Algorithm 1 by an estimate provided by (**H1**); we will hence refer to Algorithm 4 as *pseudo-marginal forward sampling* and express it concisely as

$$(\xi_{n+1}^i, \omega_{n+1}^i)_{i=1}^N \sim \mathsf{pmFS}((\xi_n^i, \omega_n^i)_{i=1}^N).$$

Data: $(\xi_n^i, \omega_n^i)_{i=1}^N$
Result: $(\xi_{n+1}^i, \omega_{n+1}^i)_{i=1}^N$

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

2   draw $I_{n+1}^i \sim \mathsf{cat}(\{\vartheta_n(\xi_n^\ell)\omega_n^\ell\}_{\ell=1}^N)$;

3   draw $\xi_{n+1}^i \sim \mathbf{P}_n(\xi_n^{I_{n+1}^i}, \cdot)$;

4   draw $\zeta_{n+1}^i \sim \mathbf{R}_n(\xi_n^{I_{n+1}^i}, \xi_{n+1}^i, \cdot)$;

5   set $\omega_{n+1}^i \leftarrow \dfrac{\ell_n \langle \zeta_{n+1}^i \rangle (\xi_n^{I_{n+1}^i}, \xi_{n+1}^i)}{\vartheta_n(\xi_n^{I_{n+1}^i}) p_n(\xi_n^{I_{n+1}^i}, \xi_{n+1}^i)}$;

Algorithm 4: Pseudo-marginal forward sampling, pmFS.

Iterating recursively, after initialisation as in Section 2.2.1, pseudo-marginal forward sampling yields a generalisation of the random-weight particle filter proposed in [17] in the context of partially observed diffusion processes.

3.2.2. Pseudo-marginal backward sampling

Let us turn our focus to backward sampling. As intractability of ℓ_n implies intractability of the kernel $\mathbf{\Pi}_n$ (defined in (2.15)), we aim at further pseudo

marginalisation by embedding $\mathbf{\Pi}_n$ into the extended kernel

$$\bar{\mathbf{\Pi}}_n(i, j, dz) \propto \omega_n^j \ell_n \langle z \rangle (\xi_n^j, \xi_{n+1}^i) \, \mathbf{R}_n(\xi_n^j, \xi_{n+1}^i, dz)$$

on $[\![1, N]\!] \times \mathcal{P}([\![1, N]\!]) \otimes \mathcal{Z}_{n+1}$. For every i, the marginal of $\bar{\mathbf{\Pi}}_n(i, \cdot)$ with respect to the j component is, by (3.5), proportional to $\omega_n^j \ell_n^\varepsilon(\xi_n^j, \xi_{n+1}^i)$, a distribution that we expect to be close to $\mathbf{\Pi}_n(i, \cdot)$ for small ε. The intractable sampling step on Line 3 in Algorithm 2 can therefore be replaced by sampling from $\bar{\mathbf{\Pi}}_n(i, \cdot)$, after which the auxiliary variables are discarded. The latter sampling operation will be examined in detail in the next section. This approach, which we express concisely as

$$(\tau_{n+1}^i)_{i=1}^N \sim \mathsf{pmBS}((\xi_n^i, \tau_n^i, \omega_n^i)_{i=1}^N, (\xi_{n+1}^i)_{i=1}^N),$$

is summarised in Algorithm 5.

Data: $(\xi_n^i, \tau_n^i, \omega_n^i)_{i=1}^N$, $(\xi_{n+1}^i)_{i=1}^N$
Result: $(\tau_{n+1}^i)_{i=1}^N$
1 **for** $i = 1 \to N$ **do**
2   **for** $j = 1 \to M$ **do**
3   draw $(J_{n+1}^{(i,j)}, \zeta_{n+1}^{(i,j)}) \sim \bar{\mathbf{\Pi}}_n(i, \cdot)$;
4   set $\tau_{n+1}^i \leftarrow \frac{1}{M} \sum_{j=1}^M \left(\tau_n^{J_{n+1}^{(i,j)}} + \tilde{h}_n(\xi_n^{J_{n+1}^{(i,j)}}, \xi_{n+1}^i) \right)$;

Algorithm 5: Pseudo-marginal backward sampling, pmBS.

It remains to discuss how to sample from the extended distribution $\bar{\mathbf{\Pi}}_n(i, \cdot)$. In the following we propose two possible approaches, which can be viewed as pseudo-marginal versions of the techniques discussed in Section 2.2.2.

Rejection sampling from $\bar{\mathbf{\Pi}}_n$

Assume that there exists some measurable nonnegative function c on X_{n+1} such that for all $(x_{n:n+1}, z) \in \mathsf{X}_n \times \mathsf{X}_{n+1} \times \mathsf{Z}_{n+1}$, $\ell_n \langle z \rangle (x_n, x_{n+1}) \leq c(x_{n+1})$. Since this condition allows the Radon–Nikodym derivative of $\bar{\mathbf{\Pi}}_n(i, \cdot)$ with respect to the probability measure

$$\rho_n^i(j, dz) \propto \omega_n^j \mathbf{R}(\xi_n^j, \xi_{n+1}^i, dz) \tag{3.7}$$

on $\mathcal{P}([\![1, N]\!]) \otimes \mathcal{Z}_{n+1}$ to be bounded uniformly as

$$\frac{d\bar{\mathbf{\Pi}}_n(i, \cdot)}{d\rho_n^i}(j, z) \leq \frac{c(\xi_{n+1}^i)\Omega_n}{\sum_{i'=1}^N \omega_n^{i'} \ell_n^\varepsilon(\xi_n^{i'}, \xi_{n+1}^i)},$$

we may sample from the target $\bar{\mathbf{\Pi}}_n(i, \cdot)$ using rejection sampling. Thus, the following procedure is iterated until acceptance: simulate a candidate (J^*, ζ^*)

from ρ_n^i by drawing $J^* \sim \mathsf{cat}(\{\omega_n^\ell\}_{\ell=1}^N)$ and $\zeta^* \sim \mathbf{R}(\xi_n^{J^*}, \xi_{n+1}^i, dz)$; then accept the same with (tractable) probability

$$\bar{\alpha}^{\mathsf{R}} := \frac{\ell_n \langle \zeta^* \rangle(\xi_n^{J^*}, \xi_{n+1}^i)}{c(\xi_{n+1}^i)}.$$

Then conditionally to acceptance, the candidate has distribution $\bar{\mathbf{\Pi}}_n(i, \cdot)$. Notably, the probability $\bar{\alpha}^{\mathsf{R}}$ is obtained by simply plugging a transition density estimate provided by (**H1**) into the probability α^{R} (see (2.17)) corresponding to the case where ℓ_n is known. Moreover, since the proposal density is independent of i, the expected complexity of this sampling schedule is linear in the number of particles (we refer again to [13]).

MCMC sampling from $\bar{\mathbf{\Pi}}_n$

In some cases, bounding the estimator $\ell_n \langle z \rangle(x_n, x_{n+1})$ uniformly in z and x_n is not possible. Still, we may sample from $\bar{\mathbf{\Pi}}_n(i, \cdot)$ using the Metropolis-Hastings algorithm with ρ_n (in (3.7)) as independent proposal. In this case, $(J_{n+1}^{(i,j)}, \zeta_{n+1}^{(i,j)})_{j=1}^M$ is a Markov chain generated recursively by the following mechanism. Given a state $J_{n+1}^{(i,j)} = J$ and $\zeta_{n+1}^{(i,j)} = \zeta$, a candidate (J^*, ζ^*) for the next state is drawn from ρ_n (as described above) and accepted with probability

$$\bar{\alpha}^{\mathsf{MH}} := 1 \wedge \frac{\ell_n \langle \zeta^* \rangle(\xi_n^{J^*}, \xi_{n+1}^i)}{\ell_n \langle \zeta \rangle(\xi_n^J, \xi_{n+1}^i)}.$$

In the case of rejection, the next state is assigned the previous state. The resulting Markov chain has $\bar{\mathbf{\Pi}}_n(i, \cdot)$ as stationary distribution and similar to the case of rejection sampling, the acceptance probability $\bar{\alpha}^{\mathsf{MH}}$ can be viewed as a plug-in estimate of the corresponding probability α^{MH} (see (2.18)).

3.2.3. Pseudo-marginal PaRIS: full update

Combining the pseudo-marginal forward and backward sampling operations yields a pseudo-marginal PaRIS update described in the following algorithm, which is the main contribution of this section.

Data: $(\xi_n^i, \tau_n^i, \omega_n^i)_{i=1}^N$
Result: $(\xi_{n+1}^i, \tau_{n+1}^i, \omega_{n+1}^i)_{i=1}^N$
1 run $(\xi_{n+1}^i, \omega_{n+1}^i)_{i=1}^N \sim \mathsf{pmFS}((\xi_n^i, \omega_n^i)_{i=1}^N)$;
2 run $(\tau_{n+1}^i)_{i=1}^N \sim \mathsf{pmBS}((\xi_n^i, \tau_n^i, \omega_n^i)_{i=1}^N, (\xi_{n+1}^i)_{i=1}^N)$;

Algorithm 6: Full pseudo-marginal PaRIS update.

Algorithm 6 is initialised by drawing $(\xi_0^i)_{i=1}^N \sim \chi^{\otimes N}$ and letting $\omega_0^i = d\chi/d\nu(\xi_0^i)$ and $\tau_n^i = 0$.

We now illustrate (**H1**) by a few examples.

Example 2 (Durham–Gallant estimator). As an illustration, we return to the state-space model framework discussed in Example 1. Let $\mathsf{X} := \mathbb{R}^{d_x}$ and $\mathsf{Y} := \mathbb{R}^{d_y}$ be equipped with their respective Borel σ-fields $\mathcal{X}$ and $\mathcal{Y}$, and let $(X_t)_{t>0}$ be some diffusion process on X driven by the homogeneous stochastic differential equation

$$dX_t = \mu(X_t)\,dt + \sigma(X_t)\,dW_t, \quad t > 0, \tag{3.8}$$

where $X_0 = 0$, $(W_t)_{t>0}$ is d_x-dimensional Brownian motion, $b : \mathsf{X} \to \mathsf{X}$ and $\sigma : \mathsf{X} \to \mathsf{X}^2$ are twice differentiable with bounded first and second order derivatives. In addition, the matrix $\sigma\sigma^\mathsf{T}$ is uniformly non-degenerate. Let $(\mathcal{F}_t)_{t>0}$ be the natural filtration generated by the process $(X_t)_{t>0}$. The state sequence $(X_t)_{t>0}$ is latent but partially observed at discrete time points $(t_n)_{n\in\mathbb{N}_{>0}}$ which are assumed to be equally spaced for simplicity, *i.e.*, $t_n = t_1 + \delta(n-1)$ for all n and some $\delta > 0$. Abusing notations, we denote $X_n := X_{t_n}$ and let q_δ be the transition density of $(X_n)_{n\in\mathbb{N}_{>0}}$. Denote by $\mathbf{Q}_\delta$ the transition kernel induced by q_δ. In general, q_δ is intractable, which makes the problem of computing online, for a given data stream $(y_n)_{n\in\mathbb{N}_{>0}}$ in Y, the sequence of joint-smoothing distributions (2.5) in models of this sort very challenging. Still, using the Euler scheme, one may, for small δ, approximate q_δ by

$$\bar{q}_\delta(x_n, x_{n+1}) := \phi(x_{n+1}; x_n + \delta\mu(x_n), \delta\sigma^2(x_n)),$$

where $\phi(\cdot; m, s^2)$ is the density of the Gaussian distribution with mean m and variance s^2. Let $\bar{\mathbf{Q}}_\delta$ be the transition kernel induced by $\bar{q}_\delta$. Since the approximation $\bar{q}_\delta$ is poor for δ not small enough, we may instead, as suggested in [16], pick some finer step size $\varepsilon \in \mathcal{E}_\delta := \{\delta/n : n \in \mathbb{N}_{>0}\}$ and estimate the density $q_\delta(x_n, x_{n+1})$ by $q_\delta\langle\zeta\rangle(x_n, x_{n+1})$, where $\zeta = (\zeta^i)_{i=1}^L$ are independent draws from some proposal $r(x_n, x_{n+1}, z)\,dz$ on $\mathsf{X}^2 \times \mathcal{X}^{\delta/\varepsilon-1}$ and

$$q_\delta\langle z\rangle(x_n, x_{n+1}) := \frac{1}{L}\sum_{i=1}^L \frac{\prod_{k=1}^{\delta/\varepsilon} \bar{q}_\varepsilon(z_{k-1}^i, z_k^i)}{r(x_n, x_{n+1}, z^i)},$$

with $z^i = (z_1^i, \ldots, z_{\delta/\varepsilon-1}^i)$ and, by convention, $z_0^i = x_n$ and $z_{\delta/\varepsilon}^i = x_{n+1}$. In [16], $r(x_n, x_{n+1}, z)\,dz$ is the distribution of a discretised (possibly modified) *Brownian bridge*, *i.e.*, Brownian motion started at x_n and conditioned to terminate at x_{n+1}.

Let Y_n denote the Y-valued observation at time t_n. We will consider two different models for the observation process $(Y_n)_{n\in\mathbb{N}}$.

Case 1

First, assume that for all $n \in \mathbb{N}_{>0}$,

$$Y_n \mid \mathcal{F}_{t_n} \sim g_{n-1}(X_{n-1}, X_n, y_n)\,dy_n,$$

where g_{n-1} is some tractable transition density with respect to Lebesgue measure. In this case, (**H1**) holds with the estimator

$$\ell_n\langle z\rangle(x_n, x_{n+1}) = q_\delta\langle z\rangle(x_n, x_{n+1})g_n(x_n, x_{n+1}, y_{n+1})$$

and the instrumental kernel

$$\mathbf{R}_n(x_n, x_{n+1}, dz) = \prod_{i=1}^{L} r(x_n, x_{n+1}, z^i) \, dz^i.$$

Finally, we note that

$$\mathbf{L}_n^\varepsilon(x_n, dx_{n+1}) = \bar{\mathbf{Q}}_\varepsilon^{\delta/\varepsilon}(x_n, dx_{n+1}) \, g_n(x_n, x_{n+1}, y_{n+1}), \tag{3.9}$$

which is generally intractable.

Case 2

Alternatively, we may assume that $(Y_n)_{n \in \mathbb{N}_{>0}}$ are discrete observations of the solution to the stochastic differential equation

$$dY_t = \tilde{\mu}(X_t, Y_t) \, dt + \tilde{\sigma}(X_t, Y_t) \, d\tilde{W}_t, \quad t > 0,$$

where $Y = 0$, $(\tilde{W}_t)_{t>0}$ is d_y-dimensional Brownian motion independent of $(W_t)_{t>0}$ and $\tilde{\mu} : \mathsf{X} \times \mathsf{Y} \to \mathsf{Y}$ and $\tilde{\sigma} : \mathsf{X} \times \mathsf{Y} \to \mathsf{Y}^2$ are known functions which are twice differentiable with bounded first and second order derivatives. In addition, the matrix $\tilde{\sigma}\tilde{\sigma}^\intercal$ is uniformly non-degenerate. Denote by p_δ the transition density of $(X_t, Y_t)_{t>0}$. In this case, the joint-smoothing distributions can, for a given data stream $(y_n)_{n \in \mathbb{N}_{>0}}$, be expressed as path measures (2.2) induced by $\ell_n(x_n, x_{n+1}) = p_\delta(x_n, y_n, x_{n+1}, y_{n+1})$, $n \in \mathbb{N}$. Since also p_δ is generally intractable we subject the bivariate process to Euler discretisation, yielding the approximation

$$\begin{aligned} &\bar{p}_\delta(x_n, y_n, x_{n+1}, y_{n+1}) \\ &:= \phi\left(x_{n+1}, y_{n+1}; (x_n + \delta\mu(x_n), y_n + \delta\tilde{\mu}(x_n, y_n))^\intercal, \delta \operatorname{diag}(\sigma^2(x_n), \tilde{\sigma}^2(x_n, y_n))\right) \end{aligned} \tag{3.10}$$

of p_δ. Denote by $\bar{\mathbf{P}}_\delta$ the Markov kernel induced by $\bar{p}_\delta$. In the case of sparse observations we may improve the approximation $\bar{p}_\delta$ by picking again some finer step size $\varepsilon \in \mathcal{E}_\delta$ and computing (by swapping, in (3.9), $\bar{q}_\delta$ for $\bar{p}_\delta$ and letting $r(x_n, y_n, x_{n+1}, y_{n+1}, z) \, dz$ be the distribution of a discretised, $\mathbb{R}^{d_x+d_y}$-valued Brownian bridge started in (x_n, y_n) and conditioned to terminate in (x_{n+1}, y_{n+1})) the Durham–Gallant estimator $p_\delta\langle z \rangle$. In this case, $\ell_n\langle z \rangle(x_n, x_{n+1}) = p_\delta\langle z \rangle(x_n, y_n, x_{n+1}, y_{n+1})$, which yields

$$\mathbf{L}_n^\varepsilon(x_n, dx_{n+1}) = \bar{p}_\varepsilon^{\delta/\varepsilon}(x_n, y_n, x_{n+1}, y_{n+1}) \, dx_{n+1}, \tag{3.11}$$

where $\bar{p}_\varepsilon^{\delta/\varepsilon}$ denotes the transition density of (the δ/ε-skeleton) $\bar{\mathbf{P}}_\varepsilon^{\delta/\varepsilon}$.

Example 3 (the exact algorithm). We consider again the partially observed diffusion process model in Example 2. In the special case where the diffusion process governed by (3.8) can be transformed into one with a constant diffusion

term through the *Lamperti transformation*, it was shown in [4, 17] how unbiased estimation of q_δ can be achieved using generalised Poisson estimators. In our setup, this simply yields $\ell_n^\varepsilon = \ell_n$ for all n. We refer to the mentioned papers for details.

Example 4 (approximate Bayesian computation smoothing). Consider smoothing in a fully dominated general state-space HMM for which the state likelihood functions $(g_n(\cdot, y_n))_{n \in \mathbb{N}}$ are intractable (or expensive to evaluate) for any given sequence $(y_n)_{n \in \mathbb{N}}$ of observations in $\mathbb{R}^{d_y}$. In the case where it is possible (or faster) to sample observation emissions according to the kernels $(\mathbf{G}_n)_{n \in \mathbb{N}}$ one may then take an *approximate-Bayesian-computation* (ABC) approach (see *e.g.* [23]), and replace any value $g_n(x_n, y_n)$ by a point estimate $\kappa_\varepsilon(\zeta_n - y_n)$, where $\zeta_n \sim \mathbf{G}_n(x_n, \cdot)$ and κ_ε is a d_y-dimensional kernel density scaled by some bandwidth $\varepsilon > 0$. In [24], the authors apply the forward-only smoothing approach of [10] to this approximate model, yielding a particle-based ABC smoothing algorithm. Also this framework is covered by (**H1**), by letting $\ell_n \langle z \rangle (x_n, x_{n+1}) = q_n(x_n, x_{n+1}) \kappa_\varepsilon(z - y_{n+1})$ and $\mathbf{R}_n(x_n, x_{n+1}, dz) = \mathbf{G}_n(x_{n+1}, dz)$. In this case,

$$\mathbf{L}_n^\varepsilon(x_n, dx_{n+1}) = \mathbf{Q}_n(x_n, dx_{n+1}) \int \kappa_\varepsilon(z - y_{n+1}) \mathbf{G}_n(x_{n+1}, dz). \qquad (3.12)$$

4. Theoretical results

4.1. Convergence of pseudo-marginal PaRIS estimates

4.1.1. Convergence of Algorithm 6

In [28], the authors established strong consistency and asymptotic normality of the PaRIS in the framework of fully dominated general state-space HMMs and the bootstrap particle filter, *i.e.*, in the simple case where $\vartheta_n \equiv 1$ and $p_n \equiv q_n$ for all n. In the following we will extend these results to the considerably more general setting comprising models (1.1) and the pseudo-marginal PaRIS in Algorithm 6. More precisely, we will show that each weighted sample $(\xi_n^i, \tau_n^i, \omega_n^i)_{i=1}^N$, $n \in \mathbb{N}$, produced by Algorithm 6 satisfies exponential concentration (Theorem 4.1) and asymptotic normality (Theorem 4.2) with respect to the expected additive functional h_n under the *skew* path model

$$\phi_{0:n}^\varepsilon(dx_{0:n}) \propto \chi(dx_0) \prod_{m=0}^{n-1} \mathbf{L}_m^\varepsilon(x_m, dx_{m+1}), \quad n \in \mathbb{N}.$$

Even though these results are established along the lines of the corresponding proofs in [28], it is the matter of non-trivial adaptations. As explained in Section 1, a random-weight particle filter (iterated pseudo-marginal forward sampling) can be viewed as a standard particle filter evolving on an extended state space comprising also the states of the auxiliary variables, and hence its convergence follows from standard SMC convergence results [17]. On the contrary, since Algorithm 6 involves *two* levels of pseudo marginalisation (with respect to

both the forward-sampling and the backward-sampling operations), it cannot be described equivalently as a special instance of the original PaRIS (even in its general form given by Algorithm 3) operating on an extended space. Thus, there is no free lunch when it concerns the theoretical analysis of this scheme. Furthermore, in the case of fully dominated HMMs and when forward sampling is guided by the bootstrap filter, which was the setting in [28], the conditional distribution of the particles given their ancestors (*i.e.*, the marginal of ρ_n in (2.8) with respect to x) coincides, at any time point n, with the denominator of the backward kernel. Mathematically, this enables a cancellation that facilitates the analysis significantly. However, this simplification is not possible once the particle dynamics is guided by general proposal kernels, which is necessarily the case in Algorithm 6. This complicates the analysis; see Remark C.4 for further details.

To be able to describe our results in full detail, we also need to introduce, for every $n \in \mathbb{N}$, the skew backward Markov kernels

$$\mathbf{B}_n^\varepsilon(x_{n+1}, dx_n) := \frac{\phi_n^\varepsilon(dx_n)\, \ell_n^\varepsilon(x_n, x_{n+1})}{\phi_n^\varepsilon[\ell_n^\varepsilon(\cdot, x_{n+1})]}$$

on $\mathsf{X}_{n+1} \times \mathcal{X}_n$ as well as the joint law

$$\mathbf{T}_n^\varepsilon(x_n, dx_{0:n-1}) := \prod_{m=0}^{n-1} \mathbf{B}_m^\varepsilon(x_{m+1}, dx_m) \tag{4.1}$$

on $\mathsf{X}_n \times \mathcal{X}^{n-1}$.

The analysis will be carried through under the following assumption.

(**H2**) For all $n \in \mathbb{N}$, the functions ϑ_n and

$$w_n : \mathsf{X}_n \times \mathsf{X}_{n+1} \times \mathsf{Z}_{n+1} \ni (x_n, x_{n+1}, z) \mapsto \frac{\ell_n \langle z \rangle(x_n, x_{n+1})}{\vartheta_n(x)p_n(x_n, x_{n+1})},$$

$$w_n^\varepsilon : \mathsf{X}_n \times \mathsf{X}_{n+1} \ni (x_n, x_{n+1}) \mapsto \frac{\ell_n^\varepsilon(x_n, x_{n+1})}{\vartheta_n(x_n)p_n(x_n, x_{n+1})}$$

are bounded. So is also $w_{-1} : \mathsf{X}_0 \ni x_0 \mapsto d\chi/d\nu(x_0)$.

Hoeffding inequality

Our first theoretical result is the following Hoeffding-type concentration inequality, which also plays a critical role for the derivation of the central limit theorem (CLT) in Theorem 4.2. For every $n \in \mathbb{N}$, let $\mathsf{A}(\mathcal{X}^n)$ denote the set of additive functionals (1.2) such that $\tilde{h}_m \in \mathsf{F}(\mathcal{X}_m \otimes \mathcal{X}_{m+1})$ for all $m \in [\![0, n-1]\!]$.

Theorem 4.1. *Assume (**H1**–2). Then for every $n \in \mathbb{N}$, $\varepsilon \in \mathcal{E}$, $h_n \in \mathsf{A}(\mathcal{X}^n)$, and $M \in \mathbb{N}_{>0}$ there exists $(c_n, d_n) \in \mathbb{R}_{>0}^2$ such that for all $N \in \mathbb{N}_{>0}$ and $\epsilon > 0$,*

$$\mathbb{P}\left(\left|\sum_{i=1}^N \frac{\omega_n^i}{\Omega_n}\tau_n^i - \phi_{0:n}^\varepsilon h_n\right| \geq \epsilon\right) \leq c_n \exp\left(-d_n N\epsilon^2\right).$$

The proof of Proposition 4.1, which is an adaptation of the proof of [28, Theorem 1], is presented in Section B. Since we proceed by induction and the objective functions $(h_n)_{n \in \mathbb{N}}$ are additive, it is, following [28], necessary to establish the result for estimators in the form $\sum_{i=1}^{N} \omega_n^i \{f_n(\xi_n^i)\tau_n^i + \tilde{f}_n(\xi_n^i)\}/\Omega_n$, where $(f_n, \tilde{f}_n) \in \mathsf{F}(\mathcal{X}_n)^2$. This is done in Proposition B.1, and Theorem 4.1 follows as a corollary of that result.

The previous bound describes the pointwise convergence of the estimator delivered by Algorithm 6 as N grows for n fixed, and here no attempt has been made to obtain a bound that is uniform in n. As we shall see shortly, in Theorem 4.6, the numerical stability of the algorithm can instead be established by bounding the asymptotic variance of the estimator. Finally, note that the previous bound implies that the estimator $\sum_{i=1}^{N} \omega_n^i \tau_n^i / \Omega_n$ tends $\mathbb{P}$-a.s. to $\phi_{0:n}^\varepsilon h_n$ as N tends to infinity.

Central limit theorem

We now focus on the asymptotic properties of Algorithm 6 and furnish, in Theorem 4.2 below, this scheme with a CLT. On the basis of this result, the stochastic stability of the algorithm is expressed by establishing, in Theorem 4.6, an $\mathcal{O}(n)$ bound on the asymptotic variances. In order to be able to state these results accurately, we need some additional notation. First, let for $(x_n, x_{n+1}) \in \mathsf{X}_n \times \mathsf{X}_{n+1}$,

$$
\begin{aligned}
&\varsigma_n^2(x_n, x_{n+1}) \\
&:= \frac{1}{w_n^\varepsilon(x_n, x_{n+1})} \int \{w_n(x_n, x_{n+1}, z) - w_n^\varepsilon(x_n, x_{n+1})\}^2 \, \mathbf{R}_n(x_n, x_{n+1}, dz) \quad (4.2)
\end{aligned}
$$

denote the conditional relative variance of the (random) weight ω_{n+1}^i given $\xi_n^{I_n^i} = x_n$ and $\xi_{n+1}^i = x_{n+1}$. In addition, for each $n \in \mathbb{N}$ and $m \in [\![0, n]\!]$, define the kernel

$$
\mathcal{R}_{m,n}^\varepsilon(x_m', dx_{0:n}) := \delta_{x_m'}(dx_m) \, \mathbf{T}_m^\varepsilon(x_m, dx_{0:m-1}) \prod_{\ell=m}^{n-1} \mathbf{L}_\ell^\varepsilon(x_\ell, dx_{\ell+1}) \qquad (4.3)
$$

on $\mathsf{X}_m \times \mathcal{X}^n$ as well as the centered version

$$
\bar{\mathcal{R}}_{m,n}^\varepsilon h(x_m) := \mathcal{R}_{m,n}^\varepsilon(h - \phi_{0:n}^\varepsilon h)(x_m) \qquad (4.4)
$$

on the same space.

The following is the main result of this section.

Theorem 4.2. *Assume (**H1**–**2**). Then for all $n \in \mathbb{N}$, $\varepsilon \in \mathcal{E}$, $M \in \mathbb{N}_{>0}$, and $h_n \in \mathsf{A}(\mathcal{X}^n)$,*

$$
\sqrt{N} \left(\sum_{i=1}^{N} \frac{\omega_n^i}{\Omega_n} \tau_n^i - \phi_{0:n}^\varepsilon h_n \right) \xrightarrow{\mathcal{D}} \sigma_n(h_n) Z,
$$

where Z is standard normally distributed and

$$\sigma_n^2(h_n) := \frac{\chi(w_{-1}\bar{\boldsymbol{\mathcal{R}}}_{0,n}^\varepsilon h_n)^2}{(\chi\mathbf{L}_{0,n-1}^\varepsilon\mathbb{1}_{\mathsf{X}^n})^2} + \sigma_n^2\langle(w_\ell^\varepsilon)_{\ell=0}^{n-1}\rangle(h_n) + \sigma_n^2\langle(\varsigma_\ell^2)_{\ell=0}^{n-1}\rangle(h_n) \qquad (4.5)$$

and

$$\sigma_n^2\langle(\varphi_\ell)_{\ell=0}^{n-1}\rangle(h_n) := \sum_{m=0}^{n-1} \frac{\phi_m^\varepsilon\vartheta_m\phi_m^\varepsilon\mathbf{L}_m^\varepsilon(\varphi_m[\bar{\boldsymbol{\mathcal{R}}}_{m+1,n}^\varepsilon h_n]^2)}{(\phi_m^\varepsilon\mathbf{L}_{m,n-1}^\varepsilon\mathbb{1}_{\mathsf{X}^n})^2}$$

$$+\sum_{m=0}^{n-1}\sum_{\ell=0}^m \frac{\phi_m^\varepsilon\vartheta_m\phi_\ell^\varepsilon\mathbf{L}_\ell^\varepsilon\{\mathbf{B}_\ell^\varepsilon(\mathbf{T}_\ell^\varepsilon h_\ell + \tilde{h}_\ell - \mathbf{T}_{\ell+1}^\varepsilon h_{\ell+1})^2\mathbf{L}_{\ell+1,m}^\varepsilon(\mathbf{B}_m^\varepsilon\varphi_m[\mathbf{L}_{m+1,n-1}^\varepsilon\mathbb{1}_{\mathsf{X}^n}]^2)\}}{M^{m-\ell+1}(\phi_\ell^\varepsilon\mathbf{L}_{\ell,m-1}^\varepsilon\mathbb{1}_{\mathsf{X}^m})(\phi_m^\varepsilon\mathbf{L}_{m,n-1}^\varepsilon\mathbb{1}_{\mathsf{X}^n})^2}$$

$$(4.6)$$

for any sequence $(\varphi_\ell)_{\ell\in\mathbb{N}}$ of measurable functions $\varphi_\ell : \mathsf{X}_\ell \times \mathsf{X}_{\ell+1} \to \mathbb{R}_{\geq 0}$.

Remark 4.3. The first term of (4.5) corresponds to the contribution of the initialisation step to the asymptotic variance. This term is incorrectly missing in the asymptotic-variance expressions given in [29, Theorem 3 and Corollary 5]. The last term corresponds to the additional variance induced by the estimation of $(\ell_n)_{n\in\mathbb{N}}$ regulated by (**H1**). Finally, as we shall see in Section 4.1.2, the first two terms correspond jointly to the variance of the ideal PaRIS in Algorithm 3, for which $(\ell_n)_{n\in\mathbb{N}}$ are assumed to be known and tractable.

Theorem 4.2 is established in Section C. The structure of the proof is adopted from [29] (however, we remark again that it is, as explained above, the matter of a non-trivial extension), and the CLT in Theorem 4.2 is obtained directly as a corollary of a more general CLT for estimators of form $\sum_{i=1}^N \omega_n^i\{f_n(\xi_n^i)\tau_n^i + \tilde{f}_n(\xi_n^i)\}/\Omega_n$, where $(f_n, \tilde{f}_n) \in \mathsf{F}(\mathcal{X}_n)^2$; see Theorem C.1.

4.1.2. Convergence of Algorithm 4

Our analysis provides, as a by-product, also a CLT for the random-weight particle filter obtained by iterating the forward-sampling operation in Algorithm 4; indeed, this result, which is similar to [17, Theorem 3], follows immediately by letting $f_n \equiv 0$ in Theorem C.1, and we state it below for completeness.

Proposition 4.4. *Assume* (**H1–2**) *and let* $(\xi_n^i, \omega_n^i)_{i=1}^N$, $n \in \mathbb{N}$, *be generated by Algorithm 4. Then for all $n \in \mathbb{N}$, $\varepsilon \in \mathcal{E}$, and $h \in \mathsf{F}(\mathcal{X}_n)$,*

$$\sqrt{N}\left(\sum_{i=1}^N \frac{\omega_n^i}{\Omega_n}h(\xi_n^i) - \phi_n^\varepsilon h\right) \xrightarrow{\mathcal{D}} \tilde{\sigma}_n(h)Z,$$

where Z is standard normally distributed and

$$\tilde{\sigma}_n^2(h) := \frac{\chi\{w_{-1}\mathbf{L}_0^\varepsilon\cdots\mathbf{L}_{n-1}^\varepsilon(h - \phi_n^\varepsilon h)\}^2}{(\chi\mathbf{L}_0^\varepsilon\cdots\mathbf{L}_{n-1}^\varepsilon\mathbb{1}_{\mathsf{X}_n})^2} + \tilde{\sigma}_n^2\langle(w_\ell^\varepsilon)_{\ell=0}^{n-1}\rangle(h) + \tilde{\sigma}_n^2\langle(\varsigma_\ell^2)_{\ell=0}^{n-1}\rangle(h)$$

$$(4.7)$$

and

$$\tilde{\sigma}_n^2 \langle (\varphi_\ell)_{\ell=0}^{n-1} \rangle (h) := \sum_{m=0}^{n-1} \frac{\phi_m^\varepsilon \vartheta_m \phi_m^\varepsilon \mathbf{L}_m^\varepsilon (\varphi_m [\mathbf{L}_{m+1}^\varepsilon \cdots \mathbf{L}_{n-1}^\varepsilon (h - \phi_n^\varepsilon h)]^2)}{(\phi_m^\varepsilon \mathbf{L}_m^\varepsilon \cdots \mathbf{L}_{n-1}^\varepsilon \mathbb{1}_{\mathsf{X}_n})^2} \quad (4.8)$$

for any sequence $(\varphi_\ell)_{\ell \in \mathbb{N}}$ of measurable functions $\varphi_\ell : \mathsf{X}_\ell \times \mathsf{X}_{\ell+1} \to \mathbb{R}_{\geq 0}$.

4.1.3. Convergence of Algorithm 3

Importantly, Theorem 4.2 provides, as another by-product, also a CLT for the ideal PaRIS in Algorithm 3. Indeed, in the case where every ℓ_n is tractable, we may set

$$\ell_n \langle z \rangle (x_n, x_{n+1}) = \ell_n(x_n, x_{n+1}) \quad (4.9)$$

for all z and define $\mathbf{R}_n$ arbitrarily. This implies that the relative variance (4.2) is identically zero, which eliminates the last term of (4.5). Thus, in this case the asymptotic variance is given by the first two terms of (4.5), but now induced by the original dynamics $(\mathbf{L}_n)_{n \in \mathbb{N}}$ (as (4.9) implies that also $\ell_n^\varepsilon = \ell_n$). This result is formulated in the following corollary, where we have defined, for each $n \in \mathbb{N}$ and $m \in [\![0, n]\!]$, the kernel

$$\boldsymbol{\mathcal{L}}_{m,n}(x_m', dx_{0:n}) := \delta_{x_m'}(dx_m) \, \mathbf{T}_n(x_m, dx_{0:m-1}) \prod_{\ell=m}^{n-1} \mathbf{L}_\ell(x_\ell, dx_{\ell+1}) \quad (4.10)$$

on $\mathsf{X}_m \times \mathcal{X}^n$ as well as the centered version

$$\bar{\boldsymbol{\mathcal{L}}}_{m,n} h(x_m) := \boldsymbol{\mathcal{L}}_{m,n}(h - \phi_{0:n} h)(x_m)$$

on the same space.

Corollary 4.5. *For each $n \in \mathbb{N}$, assume that the functions ϑ_n and*

$$\bar{w}_n : \mathsf{X}_n \times \mathsf{X}_{n+1} \ni (x_n, x_{n+1}) \mapsto \frac{\ell_n(x_n, x_{n+1})}{\vartheta_n(x) p_n(x_n, x_{n+1})}$$

are bounded and let $(\xi_n^i, \tau_n^i, \omega_n^i)_{i=1}^N$, $n \in \mathbb{N}$, be generated by Algorithm 3. Then for all $n \in \mathbb{N}$, $\varepsilon \in \mathcal{E}$, $M \in \mathbb{N}_{>0}$, and $h_n \in \mathsf{A}(\mathcal{X}^n)$,

$$\sqrt{N} \left(\sum_{i=1}^N \frac{\omega_n^i}{\Omega_n} \tau_n^i - \phi_{0:n} h_n \right) \xrightarrow{\mathcal{D}} \bar{\sigma}_n(h_n) Z,$$

where Z is standard normally distributed and

$$\bar{\sigma}_n^2(h_n) := \frac{\chi(w_{-1} \bar{\boldsymbol{\mathcal{L}}}_{0,n} h_n)^2}{(\chi \mathbf{L}_{0,n-1}^\varepsilon \mathbb{1}_{\mathsf{X}^n})^2} + \sum_{m=0}^{n-1} \frac{\phi_m \vartheta_m \phi_m \mathbf{L}_m (\bar{w}_m [\bar{\boldsymbol{\mathcal{L}}}_{m+1,n} h_n]^2)}{(\phi_m \mathbf{L}_{m,n-1} \mathbb{1}_{\mathsf{X}_n})^2}$$
$$+ \sum_{m=0}^{n-1} \sum_{\ell=0}^m \frac{\phi_m \vartheta_m \phi_\ell \mathbf{L}_\ell \{\mathbf{B}_\ell (\mathbf{T}_\ell h_\ell + \tilde{h}_\ell - \mathbf{T}_{\ell+1} h_{\ell+1})^2 \mathbf{L}_{\ell+1,m} (\mathbf{B}_m \bar{w}_m [\mathbf{L}_{m+1,n-1} \mathbb{1}_{\mathsf{X}^n}]^2)\}}{M^{m-\ell+1} (\phi_\ell \mathbf{L}_{\ell,m-1} \mathbb{1}_{\mathsf{X}_m})(\phi_m \mathbf{L}_{m,n-1} \mathbb{1}_{\mathsf{X}_n})^2}.$$
$$(4.11)$$

Corollary 4.5 extends [29, Corollary 5] to general models (1.1) and auxiliary-particle-filter-guided forward sampling.

4.2. Long-term stochastic stability

In this section we establish the long-term stochastic stability of Algorithm 6 by providing an $\mathcal{O}(n)$ bound on the sequence $(\sigma_n^2(h_n))_{n \in \mathbb{N}}$ for $M \geq 2$. Using $M \geq 2$ is critical, since, as noted in [28], using $M = 1$ in the PaRIS leads to a path-degeneracy phenomenon similar to that of the naive smoother described in Section 2.2.1; we refer to [28, Section 3.1] for a detailed discussion. Still, the variance bounds that we will present are of order $n(1 + 1/(M - 1))$, which means that large M do not serve to reduce the variance significantly. This is in good agreement with simulations, where $M = 2$ leads generally to good results. In addition, we shall see that our analysis, which is carried through in detail in Section E, yields, as by-products, a similar bound for the ideal PaRIS in Algorithm 3 as well as a time-uniform bound on the sequence $(\tilde{\sigma}^2(h))_{n \in \mathbb{N}}$ of asymptotic variances of the random-weight-particle-filter estimators generated by Algorithm 4. As far as we know, the latter result is the first of its kind. The analysis will be carried through under the following strong-mixing assumption, which is now classical (see, *e.g.*, [9, Chapter 4] and [7, Section 4.3]) and typically requires the state spaces $(\mathsf{X}_n)_{n \in \mathbb{N}}$ to be compact sets.

(**H3**) There exist constants $0 < \sigma_- < \sigma_+ < \infty$ such that for all $m \in \mathbb{N}$ and $(x_m, x_{m+1}) \in \mathsf{X}_m \times \mathsf{X}_{m+1}$,

$$\sigma_- \leq \ell_m(x_m, x_{m+1}) \leq \sigma_+$$

and

$$\sigma_- \leq \inf_{\varepsilon \in \mathcal{E}} \ell_m^\varepsilon(x_m, x_{m+1}), \quad \sup_{\varepsilon \in \mathcal{E}} \ell_m^\varepsilon(x_m, x_{m+1}) \leq \sigma_+.$$

Note that under (**H3**), each reference measure μ_m is finite; we may hence, without loss of generality, assume that each μ_m is a probability measure. Under (**H3**), define the mixing rate

$$\rho := 1 - \frac{\sigma_-}{\sigma_+} \tag{4.12}$$

as well as the constants

$$c(\sigma_\pm) := \frac{1}{\rho^2(1 - \rho)^5 \sigma_-}, \quad d(\sigma_\pm) := \frac{1}{(1 - \rho)^4 \sigma_-^2} \left(2 + \frac{1}{1 - \rho}\right)^2,$$

and

$$\tilde{c}(\sigma_\pm) := \frac{1}{(1 - \rho)^3 \sigma_-} \left(\frac{1}{\rho^2(1 - \rho^2)} + 1\right).$$

Having introduced these quantities, we are ready to present the main result of this section.

Theorem 4.6. *Assume (**H1**–**3**). Then for every $\varepsilon \in \mathcal{E}$, $M \geq 2$, $h_n \in \mathsf{A}(\mathcal{X}^n)$,*

and sequence $(\varphi_\ell)_{\ell \in \mathbb{N}}$ *of bounded measurable functions* $\varphi_\ell : \mathsf{X}_\ell \times \mathsf{X}_{\ell+1} \to \mathbb{R}_{\geq 0}$,

$$\limsup_{n \to \infty} \frac{1}{n} \sigma_n^2 \langle (\varphi_\ell)_{\ell=0}^{n-1} \rangle (h_n)$$
$$\leq \left(c(\sigma_\pm) + \frac{1}{M-1} d(\sigma_\pm) \right) \sup_{\ell \in \mathbb{N}} \|\tilde{h}_\ell\|_\infty^2 \sup_{\ell \in \mathbb{N}} \|\vartheta_\ell\|_\infty \sup_{\ell \in \mathbb{N}} \|\varphi_\ell\|_\infty.$$

Using Theorem 4.6, $\mathcal{O}(n)$ bounds on the asymptotic variances of Algorithm 6 and Algorithm 3 are readily obtained.

Corollary 4.7. *Assume* (**H1–3**)*. Then for every* $\varepsilon \in \mathcal{E}$, $M \geq 2$, *and* $h_n \in \mathsf{A}(\mathcal{X}^n)$,

$$\limsup_{n \to \infty} \frac{1}{n} \sigma_n^2(h_n)$$
$$\leq \left(c(\sigma_\pm) + \frac{1}{M-1} d(\sigma_\pm) \right) \sup_{\ell \in \mathbb{N}} \|\tilde{h}_\ell\|_\infty^2 \sup_{\ell \in \mathbb{N}} \|\vartheta_\ell\|_\infty \left(\sup_{\ell \in \mathbb{N}} \|w_\ell^\varepsilon\|_\infty + \sup_{\ell \in \mathbb{N}} \|\varsigma_\ell^2\|_\infty \right)$$
$$+ \frac{1}{\rho^2 (1-\rho)^4 (\chi \mathbb{1}_{\mathsf{X}_0})^2} \sup_{\ell \in \mathbb{N}} \|\tilde{h}_\ell\|_\infty^2. \quad (4.13)$$

Corollary 4.8. *Assume* (**H1–3**)*. Then for every* $M \geq 2$ *and* $h_n \in \mathsf{A}(\mathcal{X}^n)$,

$$\limsup_{n \to \infty} \frac{1}{n} \bar{\sigma}_n^2(h_n) \leq \left(c(\sigma_\pm) + \frac{1}{M-1} d(\sigma_\pm) \right) \sup_{\ell \in \mathbb{N}} \|\tilde{h}_\ell\|_\infty^2 \sup_{\ell \in \mathbb{N}} \|\vartheta_\ell\|_\infty \sup_{\ell \in \mathbb{N}} \|\bar{w}_\ell\|_\infty$$
$$+ \frac{1}{\rho^2 (1-\rho)^4 (\chi \mathbb{1}_{\mathsf{X}_0})^2} \sup_{\ell \in \mathbb{N}} \|\tilde{h}_\ell\|_\infty^2,$$

where $(\bar{\sigma}_n^2)_{n \in \mathbb{N}}$ *is given by* (4.11).

Finally, we provide, for completeness, a time-uniform bound on the asymptotic variances of the random-weight particle filter corresponding to repeated forward sampling (Algorithm 4). This bound is obtained more or less for free while establishing Theorem 4.6 (see Section E for details).

Proposition 4.9. *Assume* (**H1–2**) *and* (**H3**)*. Then for every* $\varepsilon \in \mathcal{E}$ *and* $h \in \mathsf{F}(\mathcal{X}_n)$,

$$\tilde{\sigma}_n^2(h) \leq \tilde{c}(\sigma_\pm) \|h\|_\infty^2 \sup_{\ell \in \mathbb{N}} \|\vartheta_\ell\|_\infty \left(\sup_{\ell \in \mathbb{N}} \|w_\ell^\varepsilon\|_\infty + \sup_{\ell \in \mathbb{N}} \|\varsigma_\ell^2\|_\infty \right)$$
$$+ \|h\|_\infty^2 \frac{\rho^{2n}}{\rho^4 (1-\rho)^2 (\chi \mathbb{1}_{\mathsf{X}_0})^2},$$

where $(\tilde{\sigma}_n^2)_{n \in \mathbb{N}}$ *are given by* (4.7).

4.3. Bounds on asymptotic bias

In the previous section we saw that asymptotically, as N tends to infinity, the estimator produced by n iterations of Algorithm 6 converges to the 'skew' expectation $\phi_{0:n}^\varepsilon h_n$. In this part we will study the discrepancy between $\phi_{0:n}^\varepsilon h_n$ and

$\phi_{0:n}h_n$ and establish an $\mathcal{O}(n\varepsilon)$ bound on the same. The analysis will be performed under the assumption that the precision parameter ε controls, uniformly in n, the bias of the estimators provided by (**H1**) in the following sense.

(**H4**) There exists a constant $c > 0$ such that for all $n \in \mathbb{N}$, $\varepsilon \in \mathcal{E}$, $h \in \mathsf{F}(\mathcal{X}_n \otimes \mathcal{X}_{n+1})$, and $x \in \mathsf{X}_n$,

$$|\mathbf{L}_n^{\varepsilon}h(x) - \mathbf{L}_nh(x)| \leq c\varepsilon\|h\|_{\infty}.$$

Example 5 (Durham–Gallant estimator, cont.). We check (**H4**) for the Durham–Gallant estimator in Example 2.

Case 1

Assume that the emission densities of the model are uniformly bounded, *i.e.*, there exists $\sigma_+ \in \mathbb{R}_{>0}$ such that $\|g_n\|_{\infty} \leq \sigma_+$ for all $n \in \mathbb{N}$. In the case of the Euler scheme and under the given assumptions on equation (3.8), it can be shown that there exist $c_\delta > 0$ and $d_\delta > 0$ such that for all $\varepsilon \in \mathcal{E}_\delta$ and $(x_n, x_{n+1}) \in \mathsf{X}^2$,

$$\left|q_\delta(x_n, x_{n+1}) - \int \bar{\mathbf{Q}}_\varepsilon^{\delta/\varepsilon-1}(x_n, dx)\, \bar{q}_\varepsilon(x, x_{n+1})\right| \leq c_\delta \frac{\varepsilon}{\delta} \exp\left(-d_\delta\|x_{n+1} - x_n\|^2\right); \tag{4.14}$$

see [2] (see also [11] for an application to SMC methods). Thus, using (3.9), for all x_n and $h \in \mathsf{F}(\mathcal{X}^{\otimes 2})$,

$$\begin{aligned}
|\mathbf{L}_n^{\varepsilon}&h(x_n) - \mathbf{L}_nh(x_n)| \\
&\leq c_\delta \frac{\varepsilon}{\delta} \int h(x_n, x_{n+1})g(x_n, x_{n+1}, y_{n+1}) \exp\left(-d_\delta\|x_{n+1} - x_n\|^2\right) \, dx_{n+1} \\
&\leq c_\delta \frac{\varepsilon}{\delta} \left(\frac{\pi}{d_\delta}\right)^{d_x/2} \sigma_+\|h\|_{\infty}.
\end{aligned}$$

Case 2

The second case can be treated straightforwardly by combining the bound (4.14), applied to the bivariate process $(X_t, Y_t)_{t>0}$, with (3.11). This provides the existence of constants $\tilde{c}_\delta > 0$ and $\tilde{d}_\delta > 0$ such that for all $\varepsilon \in \mathcal{E}_\delta$ and $(x_n, y_n, y_{n+1}) \in \mathsf{X} \times \mathsf{Y}^2$ and $h \in \mathsf{F}(\mathcal{X}^{\otimes 2})$,

$$\begin{aligned}
|\mathbf{L}_n^{\varepsilon}&h(x_n) - \mathbf{L}_nh(x_n)| \\
&\leq \tilde{c}_\delta \frac{\varepsilon}{\delta} \exp\left(-\tilde{d}_\delta\|y_{n+1} - y_n\|^2\right) \int h(x_n, x_{n+1}) \exp\left(-\tilde{d}_\delta\|x_{n+1} - x_n\|^2\right) \, dx_{n+1} \\
&\leq \tilde{c}_\delta \frac{\varepsilon}{\delta} \left(\frac{\pi}{\tilde{d}_\delta}\right)^{d_x/2} \|h\|_{\infty}.
\end{aligned}$$

Example 6 (the exact algorithm, cont.). In this case, the estimator is unbiased; thus, (**H4**) holds true for $\varepsilon = 0$.

Example 7 (ABC smoothing, cont.). In [24], the authors carry through their theoretical analysis under the assumption that each emission density g_n is Lipschitz in the sense that there exists some constant $L \in \mathbb{R}_{>0}$ such that

$$\sup_{x_n \in \mathsf{X}_n} |g_n(x_n, y_n) - g_n(x_n, y'_n)| \le L \|y_n - y'_n\|_1 \tag{4.15}$$

for all y_n and y'_n in $\mathbb{R}^{d_y}$. For the purpose of illustration, assume that κ_ε is a zero-mean multivariate normal distribution with covariance matrix $\varepsilon^2 \mathbf{I}_{d_y}$ for $\varepsilon > 0$. It is then easily shown that the condition (4.15) implies (**H4**); indeed, in this case, for all $x_n \in \mathsf{X}_n$ and $h \in \mathsf{F}(\mathcal{X}_n \otimes \mathcal{X}_{n+1})$, using (3.12),

$$|\mathbf{L}_n^\varepsilon h(x_n) - \mathbf{L}_n h(x_n)| \le L \|h\|_\infty \int \kappa_\varepsilon(z - y_{n+1}) \|z - y_{n+1}\|_1 \, dz \le \varepsilon d_y L \|h\|_\infty.$$

Under (**H3**) and (**H4**) we may establish the next theorem, whose proof is postponed to Section D.

Theorem 4.10. *Assume (**H1**) and (**H3–4**). Then for all $n \in \mathbb{N}$, $\varepsilon \in \mathcal{E}$, and $h_n \in \mathsf{A}(\mathcal{X}^n)$,*

$$\varepsilon^{-1} \left| \phi_{0:n}^\varepsilon h_n - \phi_{0:n} h_n \right| \le 2c \frac{\sigma_+}{\sigma_-^2} \sum_{k=0}^{n-1} \|\tilde{h}_k\|_\infty \left(\sum_{m=1}^{n-1} \rho^{|k-m|-1} + 1 \right)$$

$$\le 2cn \frac{\sigma_+}{\sigma_-^2} \left(1 + \frac{1}{\rho} + \frac{2}{1-\rho} \right) \sup_{k \in [\![0, n-1]\!]} \|\tilde{h}_k\|_\infty,$$

*where c is the constant in (**H4**).*

In the case where $\sup_{k \in \mathbb{N}} \|\tilde{h}_k\|_\infty < \infty$, the bound provided by Theorem 4.10 is $\mathcal{O}(n)$. Moreover, by letting $\tilde{h}_k \equiv 0$, for $k \in [\![0, n-2]\!]$ and $\tilde{h}_{n-1}(x_{n-1}, x_n) = h(x_n)$ for some given objective function $h \in \mathsf{F}(\mathcal{X}_n)$, Theorem 4.10 provides, as a by-product, the following uniform error bound for the marginals (a result referred to as the *filter sensitivity* in the case of parametric state-space models).

Corollary 4.11. *Assume (**H1**) and (**H3–4**). Then for all $n \in \mathbb{N}$, $\varepsilon \in \mathcal{E}$, and $h \in \mathsf{A}(\mathcal{X}^n)$,*

$$\varepsilon^{-1} \left| \phi_n^\varepsilon h - \phi_n h \right| \le 2c \frac{\sigma_+}{\sigma_-^2} \|h\|_\infty \left(1 + \frac{1}{\rho(1-\rho)} \right).$$

Remark 4.12. Consider now a parameterised version of the model, where the transition densities $(\ell_{n;\theta})_{k \in \mathbb{N}}$ are indexed by some parameter θ belonging to some parameter space Θ being a subset of $\mathbb{R}^d$. Assume further that all $(\ell_{n;\theta})_{n \in \mathbb{N}}$ are differentiable with respect to θ and such that for all $n \in \mathbb{N}$, $x_n \in \mathsf{X}$, and

$h \in \mathsf{F}(\mathcal{X}_{n+1})$,

$$\nabla_\theta \int \ell_{n;\theta}(x_n, x_{n+1}) h(x_{n+1}) \, \mu_{n+1}(dx_{n+1})$$
$$= \int \nabla_\theta \ell_{n;\theta}(x_n, x_{n+1}) h(x_{n+1}) \, \mu_{n+1}(dx_{n+1})$$

and

$$\sup_{\theta \in \Theta} \int |\nabla_\theta \ell_{n;\theta}(x_n, x_{n+1})| \, \mu_{n+1}(dx_{n+1}) \leq c < \infty$$

for some positive constant c, implying (**H4**) in the sense that for all $n \in \mathbb{N}$, $h \in \mathsf{F}(\mathcal{X}_{n+1})$, and $x_n \in \mathsf{X}$,

$$|\mathbf{L}_{n;\theta} h(x_n) - \mathbf{L}_{n;\theta'} h(x_n)| \leq c\|\theta - \theta'\| \|h\|_\infty$$

(*i.e.*, $\varepsilon = \|\theta - \theta'\|$ in this case). Finally, assume also that family satisfies (**H3**) uniformly over the parameter space in the sense that for all $n \in \mathbb{N}$, $(x_n, x_{n+1}) \in \mathsf{X}^2$, and $\theta \in \Theta$, $\sigma_- \leq \ell_{n;\theta}(x_n, x_{n+1}) \leq \sigma_+$. Then Theorem 4.10 provides a positive constant d such that for all $n \in \mathbb{N}$,

$$\left|\phi_{0:n;\theta} h_n - \phi_{0:n;\theta'} h_n\right| \leq dn\|\theta - \theta'\| \sup_{k \in \mathbb{N}} \|\tilde{h}_k\|_\infty.$$

This extends previous results on the uniform continuity of the filter distribution (see, *e.g.*, [30, 22]) to smoothing of additive state functionals.

5. A numerical example

An exhaustive numerical analysis of the methods and results presented above is beyond the scope of the present paper, and is left as future research. Some numerical illustrations of Algorithm 6 in the special case of unbiased pseudo-marginalisation ($\varepsilon = 0$) via the exact algorithm (see Example 3 and Example 6) are provided in [19]. In this section, we focus on biased estimation ($\varepsilon > 0$) and illustrate the Lipschitz continuity established by Theorem 4.10 on the basis of a simple model that allows, as a comparison, a fully analytical solution to the additive smoothing problem.

In the following we will consider an instance of Example 2, Case 1, where the latent diffusion $(X_t)_{t>0}$ is an *Ornstein–Uhlenbeck process* [36] on $\mathbb{R}$ parameterised by

$$\mu(x) = -(x - \theta), \quad \sigma(x) \equiv 1,$$

where $\theta \in \mathbb{R}$ is a parameter. This process is assumed to be initialised according to the standard normal distribution. Furthermore, conditionally to $(X_t)_{t>0}$, observations $(Y_n)_{n \in \mathbb{N}}$ are generated as

$$Y_n = (1 - \varepsilon)\varphi(X_n) + \eta_n, \quad n \in \mathbb{N}_{>0},$$

where $t_n = \delta n$ for some given observation interval $\delta > 0$, $X_n = X_{t_n}$, $(\eta_n)_{n \in \mathbb{N}_{>0}}$ are mutually independent and standard normally distributed noise variables,

$\varepsilon \in \mathcal{E} \coloneqq [0,1]$ is a parameter, and φ is a bounded measurable function on $\mathbb{R}$. We let $\mathbf{Q}_\delta$ denote the (Gaussian) Markov transition kernel of the time-discrete chain $(X_n)_{n \in \mathbb{N}}$.

In this toy example, our aim is to illustrate Theorem 4.10 by viewing models with $\varepsilon > 0$ as 'skew' versions of a 'true' model with $\varepsilon = 0$. Given simulated data $(y_n)_{n \in \mathbb{N}_{>0}}$ from the true model, smoothed additive expectations under skew models with different $\varepsilon > 0$ will be compared to the same expectations under the true model. Note that the model specified above satisfies condition (**H4**); indeed, denote by $g^\varepsilon(x_n, y_n)$ the emission density of Y_n given X_n, which is Gaussian with mean $(1 - \varepsilon)\varphi(x_n)$ and unit variance, and by $g(x_n, y_n)$ the same density for the true model. Then by the mean-value theorem, for all $\varepsilon \in \mathcal{E}$ and $(x_n, y_n) \in \mathbb{R}^2$,

$$
\begin{aligned}
|g^\varepsilon(x_n, y_n) - g(x_n, y_n)| &\leq \varepsilon \sup_{\varepsilon \in \mathcal{E}} \left| \frac{\partial}{\partial \varepsilon} g^\varepsilon(x_n, y_n) \right| \\
&\leq \varepsilon \left(|\varphi(x_n)y_n - \varphi^2(x_n)| + \varphi^2(x_n) \right),
\end{aligned}
$$

implying immediately that for all bounded measurable real-valued functions h,

$$
|\mathbf{L}_n^\varepsilon h(x_n) - \mathbf{L}_n h(x_n)| \leq \varepsilon c(y_{n+1}) \|h\|_\infty,
$$

where $\mathbf{L}_n^\varepsilon(x_n, dx_{n+1}) \coloneqq \mathbf{Q}_\delta(x_n, dx_{n+1}) g^\varepsilon(x_{n+1}, y_{n+1})$ and $c(y_{n+1}) \coloneqq \|\varphi\|_\infty |y_{n+1}| + 2\|\varphi\|_\infty^2$. Thus, (**H4**) holds true under the mild assumption that $\sup_{n \in \mathbb{N}_{>0}} |y_n| < \infty$.

In this context, we conducted numerical experiments with $\theta = 5$ and $\varphi(x) = \min(\max(x, -10^5), 10^5)$. With this parametrisation, the model is, in practice, linear and Gaussian; thus, for linear additive state functionals in the form $h_n(x_{0:n}) = \sum_{k=0}^n x_k$, a very good approximation of the exact solution $(\phi_{0:n}^\varepsilon h_n)_{n \in \mathbb{N}}$ to the optimal smoothing problem can, for the skewed models ($\varepsilon > 0$) as well as the true model ($\varepsilon = 0$), be obtained using Kalman recursions [33]. Figure 1a displays the discrepancy between $\phi_{0:n}^\varepsilon h_n$ and $\phi_{0:n} h_n$ for varying $\varepsilon \in \{0, 0.05, 0.1, \ldots, 0.5\}$ and a fixed $n = 50$ (the red line). Clearly, the bias increases—in perfect agreement with Theorem 4.10—at a rate that is at most linear in ε. In addition, Figure 1a shows similar biases (turquoise markers) obtained using the ideal PaRIS in Algorithm 3 with $N = 200$ particles and $M = 2$ backward samples. In this algorithm, the particles were propagated using the *optimal importance function* [15], which can be computed explicitly in the linear Gaussian case. Here the PaRIS was re-run 60 times for each value of ε. Note that the variance of the PaRIS estimates increases somewhat with ε, reflecting the fact that the proposal becomes less and less compatible with the data as the model gets increasingly skew.

In order to illustrate further the $\mathcal{O}(n\varepsilon)$ bound provided by Theorem 4.10 as well as the stochastic stability of Algorithm 3 established in Corollary 4.8, Kalman smoothing was conducted for $n \in [\![1, 50]\!]$ on the basis of a skew model with fixed $\varepsilon = 0.1$. Figure 1b shows, as expected, a linear increase of the bias with n (red line). In addition, displaying also the errors (green lines) of 60
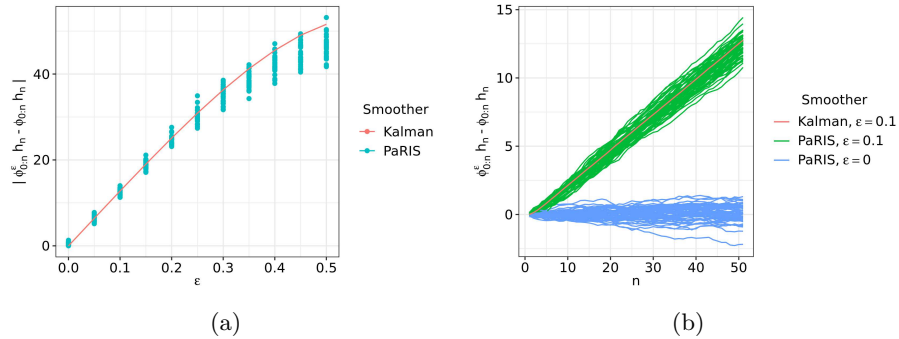
Fig 1: (a) The red line is the deviation of $\phi_{0:n}^{\varepsilon}h_n$ from $\phi_{0:n}h_n$ for varying $\varepsilon \in \{0, 0.05, 0.1, \ldots, 0.5\}$ and fixed $n = 50$, computed by means of Kalman smoothing. Turquoise markers are similar errors obtained on the basis of 60 replicates of the ideal PaRIS in Algorithm 3 with $N = 200$ particles and $M = 2$ backward samples. (b) The red line is the same deviation for increasing $n \in [\![1, 50]\!]$ and fixed $\varepsilon = 0.1$. Then green and blue lines correspond to errors of 60 independent ideal PaRIS replicates obtained under the skew and true models, respectively, for $N = 200$ and $M = 2$.

independent ideal PaRIS replicates (obtained under the skew model with the same algorithmic parametrisation as previously), the same plot confirms the stochastic stability of the ideal PaRIS algorithm, as the variance does not grow faster than linearly with n. Finally, for completeness Figure 1b reports similar errors (blue lines) when the ideal PaRIS evolves under the dynamics of the true model, and in this case the bias is negligible as expected (whereas the variance is still increasing linearly).

Acknowledgements

References

[1] C. Andrieu and G. O. Roberts. The pseudo-marginal approach for efficient Monte Carlo computations. *Ann. Statist.*, 37(2):697–725, 2009.

[2] V. Bally and D. Talay. The law of the Euler scheme for stochastic differential equations: II. Approximation of the density. *Monte Carlo Methods and Applications*, 2:93–128, 1996.

[3] M. A. Beaumont. Estimation of population growth or decline in genetically monitored populations. *Genetics*, 164(3):1139–1160, 2003.

[4] A. Beskos, O. Papaspiliopoulos, G.O. Roberts, and P. Fearnhead. Exact and computationally efficient likelihood-based estimation for discretely observed diffusion processes. *J. R. Stat. Soc. Ser. B*, 68(3):333–382, 2006. With discussions.

[5] L. M. Candanedo, V. Feldheim, and D. Deramaix. A methodology based on hidden Markov models for occupancy detection and a case study in a low energy residential building. *Energy and Buildings*, 148:327–341, 2017.

[6] O. Cappé. Online EM algorithm for hidden Markov models. *J. Comput. Graph. Statist.*, 20(3):728–749, 2011.

[7] O. Cappé, É. Moulines, and T. Ryden. *Inference in Hidden Markov Models*. Springer, 2005.

[8] N. Chopin and O. Papaspiliopoulos. *An Introduction to Sequential Monte Carlo Methods*. Springer, 2020.

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

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

[11] P. Del Moral, J. Jacod, and P. Protter. The Monte-Carlo method for filtering with discrete-time observations. *Probab. Theory Relat. Fields*, 123:346–368, 2001.

[12] P. Del Moral, E. Moulines, J. Olsson, and C. Vergé. Convergence properties of weighted particle islands with application to the double bootstrap algorithm. *Stochastic Systems*, 2(6):367–419, 2016.

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

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

[15] A. Doucet, S. Godsill, and C. Andrieu. On sequential Monte Carlo sampling methods for Bayesian filtering. *Statistics and computing*, 10(3):197–208, 2000.

[16] G. B. Durham and A. R. Gallant. Numerical techniques for maximum likelihood estimation of continuous-time diffusion processes. *Journal of Business & Economic Statistics*, 20(3):297–316, 2002.

[17] P. Fearnhead, O. Papaspiliopoulos, and G. O. Roberts. Particle filters for partially observed diffusions. *J. Roy. Statist. Soc. B*, 70(4):755–777, 2008.

[18] P. Fearnhead, O. Papaspiliopoulos, G. O. Roberts, and A. Stuart. Random weight particle filtering of continuous time processes. *J. Roy. Statist. Soc. B*, 72(4):497–512, 2010.

[19] P. Gloaguen, M.-P. Etienne, and S. Le Corff. Online sequential Monte Carlo smoother for partially observed diffusion processes. *EURASIP Journal on Advances in Signal Processing*, 2018(1):9, 2018.

[20] S. J. Godsill, A. Doucet, and M. West. Monte Carlo smoothing for nonlinear time series. *J. Amer. Statist. Assoc.*, 50:438–449, 2004.

[21] F. Le Gland and L. Mevel. Recursive estimation in HMMs. In *Proc. IEEE*

Conf. Decis. Control, pages 3468–3473, 1997.

[22] F. Le Gland and N. Oudjane. Stability and uniform approximation of nonlinear filters using the Hilbert metric and application to particle filters. *Ann. Appl. Probab.*, 14:144–187, 2004.

[23] J. M. Marin, P. Pudlo, C. P. Robert, and R. J. Ryder. Approximate Bayesian computational methods. *Stat. Comput.*, 22:1167–1180, 2012.

[24] J. S. Martin, A. Jasra, S. S. Singh, N. Whiteley, P. Del Moral, and E. Mc-Coy. Approximate Bayesian Computation for smoothing. *Stochastic Analysis and Applications*, 32:397–420, 1998.

[25] J. M. Mcgree, C. C. Drovandi, G. White, and A. N. Pettitt. A pseudo-marginal sequential Monte Carlo algorithm for random effects models in Bayesian sequential design. *Stat. Comput.*, 26:1121–1136, 2016.

[26] T. Michelot, R. Langrock, and T. A. Patterson. moveHMM: an R package for the statistical modelling of animal movement data using hidden Markov models. *Methods in Ecology and Evolution*, 7(11):1308–1315, 2016.

[27] J. Olsson and J. Ströjby. Particle-based likelihood inference in partially observed diffusion processes using generalised Poisson estimators. *Electron. J. Statist.*, 5:1090–1122, 2011.

[28] J. Olsson and J. Westerborn. Efficient particle-based online smoothing in general hidden Markov models: The PaRIS algorithm. *Bernoulli*, 23(3):1951–1996, 2017.

[29] J. Olsson and J. Westerborn Alenlöv. Particle-based online estimation of tangent filters with application to parameter estimation in nonlinear state-space models. *Annals of the Institute of Statistical Mathematics*, 72(2):545–576, 2018.

[30] A. Papavasiliou. Parameter estimation and asymptotic stability in stochastic filtering. *Stoch. Proc. Appl.*, 116(7):1048–1065, 2006.

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

[32] L. R. Rabiner. A tutorial on hidden Markov models and selected applications in speech recognition. In *Proceedings of the IEEE*, pages 257–286, 1989.

[33] H. E. Rauch, F. Tung, and C. T. Striebel. Maximum likelihood estimates of linear dynamic systems. *AIAA journal*, 3(8):1445–1450, 1965.

[34] S. Särkkä. *Bayesian Filtering and Smoothing*. Cambridge University Press, New York, 2013.

[35] S. Särkkä, A. Vehtari, and J. Lampinen. Rao-Blackwellized particle filter for multiple target tracking. *Information Fusion*, 8(1):2–15, 2007.

[36] G. E Uhlenbeck and L. S Ornstein. On the theory of the Brownian motion. *Phys. rev.*, 36(5):823, 1930.

[37] C. Yau, O. Papaspiliopoulos, G. O. Roberts, and C. Holmes. Bayesian nonparametric hidden Markov models with applications in genomics. 73:1–21, 2011.

[38] S. Yonekura and A. Beskos. Online smoothing for diffusion processes observed with noise, 2020. ArXiv:2003.12247.

[39] W. Zucchini, I. L. Mac Donald, and R. Langrock. *Hidden Markov Models*

for Time Series: an Introduction using R. CRC Press, 2017.

Appendix A: Proof of Lemma 2.2

To show (i), write, using the definition (2.10), the right-hand side as

$$
\iint \phi_n \mathbf{L}_n(dx_{n+1}) \, \mathbf{B}_n(x_{n+1}, dx_n) \, h(x_n, x_{n+1})
$$
$$
= \iint \phi_n[\ell_n(\cdot, x_{n+1})] \, \mu_{n+1}(dx_{n+1}) \frac{\phi_n(dx_n)[\ell_n(\cdot, x_{n+1})]}{\phi_n[\ell_n(\cdot, x_{n+1})]} \, h(x_n, x_{n+1})
$$
$$
= \iint \phi_n(dx_n) \, \mathbf{L}_n(x_n, dx_{n+1}) \, h(x_n, x_{n+1}),
$$

which was to be established.

The statement (ii) is shown by induction. Thus, assume that the claim holds true for n, pick arbitrarily $h \in \mathsf{F}(\mathcal{X}^{n+1})$, and write, using the induction hypothesis,

$$
\phi_{0:n} \mathbf{L}_{0,n} h = \iint \phi_n(dx_n) \, \mathbf{T}_n(x_n, dx_{0:n-1}) \, \mathbf{L}_n(x_n, dx_{n+1}) \, h(x_{0:n+1})
$$
$$
= \iint \phi_n(dx_n) \, \mathbf{L}_n(x_n, dx_{n+1}) \, \bar{h}(x_n, x_{n+1}),
$$

where

$$
\bar{h}(x_n, x_{n+1}) \coloneqq \int \mathbf{T}_n(x_n, dx_{0:n-1}) h(x_{0:n+1}).
$$

Now, since $\mathbf{B}_n \bar{h}(x_{n+1}) = \mathbf{T}_{n+1} h(x_{n+1})$, (i) yields

$$
\phi_{0:n} \mathbf{L}_{0,n} h = \iint \phi_n \mathbf{L}_n(dx_{n+1}) \, \mathbf{B}_n(x_{n+1}, dx_n) \, \bar{h}(x_n, x_{n+1})
$$
$$
= \int \phi_n \mathbf{L}_n(dx_{n+1}) \, \mathbf{T}_{n+1} h(x_{n+1}).
$$

Thus, we obtain, by applying (2.3) and (2.4),

$$
\phi_{0:n+1} h = \frac{\phi_{0:n} \mathbf{L}_{0,n} h}{\phi_{0:n} \mathbf{L}_{0,n} \mathbb{1}_{\mathsf{X}^n}} = \int \frac{\phi_n \mathbf{L}_n(dx_{n+1})}{\phi_n \mathbf{L}_n \mathbb{1}_{\mathsf{X}_{n+1}}} \mathbf{T}_{n+1} h(x_{n+1}) = \phi_{n+1} \mathbf{T}_{n+1} h.
$$

Finally, we note that the base case $n = 1$ follows straightforwardly by combining (i) and (2.4).

Appendix B: Proof of Theorem 4.1

As explained in Section 4.1.1, Theorem 4.1 follows as an immediate corollary of the following more general result, to whose proof we devote this section.

Proposition B.1. *Assume (**H**1–2). Then for all $n \in \mathbb{N}$, $h_n \in \mathsf{A}(\mathcal{X}^n)$, $(f_n, \tilde{f}_n) \in$ $\mathsf{F}(\mathcal{X}_n)^2$, and $M \in \mathbb{N}_{>0}$ there exist $(c_n, d_n) \in \mathbb{R}_{>0}^2$ such that for all $N \in \mathbb{N}_{>0}$ and $\epsilon > 0$,*

$$
(i) \quad \mathbb{P}\left(\left| \frac{1}{N} \sum_{i=1}^{N} \omega_n^i \{ \tau_n^i f_n(\xi_n^i) + \tilde{f}_n(\xi_n^i) \} - \frac{\phi_{n-1}^\varepsilon \mathbf{L}_{n-1}^\varepsilon (\mathbf{T}_n^\varepsilon h_n f_n + \tilde{f}_n)}{\phi_{n-1}^\varepsilon \vartheta_{n-1}} \right| \geq \epsilon \right)
$$

$$
\leq c_n \exp\left(-d_n N \epsilon^2 \right),
$$

$$
(ii) \quad \mathbb{P}\left(\left| \sum_{i=1}^{N} \frac{\omega_n^i}{\Omega_n} \{ \tau_n^i f_n(\xi_n^i) + \tilde{f}_n(\xi_n^i) \} - \phi_n^\varepsilon (\mathbf{T}_n^\varepsilon h_n f_n + \tilde{f}_n) \right| \geq \epsilon \right)
$$

$$
\leq c_n \exp\left(-d_n N \epsilon^2 \right).
$$

Note that letting $f_n \equiv 1$ and $\tilde{f}_n \equiv 0$ in (ii) yields immediately Theorem 4.1.

We preface the proof of Proposition B.1 by a technical lemma, generalising [28, Lemma 12] and [19, Lemma 2], which will be instrumental in the following. For all $n \in \mathbb{N}$, define the following σ-fields:

$$
\mathcal{F}_n^N := \sigma\{ (\xi_m^i, \omega_m^i, \tau_m^i) : i \in [\![1, N]\!], m \in [\![0, n]\!] \},
$$

$$
\mathcal{G}_{n+1}^N := \sigma\{ (\xi_{n+1}^i, I_{n+1}^i) : i \in [\![1, N]\!] \} \vee \mathcal{F}_n^N.
$$

Lemma B.2. *Assume (**H**1–2). For all $n \in \mathbb{N}$, $(f_{n+1}, \tilde{f}_{n+1}) \in \mathsf{F}(\mathcal{X}_{n+1})^2$, and $(N, M) \in \mathbb{N}_{>0}^2$, the random variables $\omega_{n+1}^i \{ \tau_{n+1}^i f_{n+1}(\xi_{n+1}^i) + \tilde{f}_{n+1}(\xi_{n+1}^i) \}$, $i \in [\![1, N]\!]$, are independent and identically distributed (i.i.d.) conditionally to $\mathcal{F}_n^N$. In addition, for all i,*

$$
\mathbb{E}\left[\omega_{n+1}^i \{ \tau_{n+1}^i f_{n+1}(\xi_{n+1}^i) + \tilde{f}_{n+1}(\xi_{n+1}^i) \} \mid \mathcal{F}_n^N \right]
$$

$$
= (\phi_n^N \vartheta_n)^{-1} \sum_{j=1}^{N} \frac{\omega_n^j}{\Omega_n} \{ \tau_n^j \mathbf{L}_n^\varepsilon f_{n+1}(\xi_n^j) + \mathbf{L}_n^\varepsilon (\tilde{h}_n f_{n+1} + \tilde{f}_{n+1})(\xi_n^j) \}.
$$

Proof. Independence and equal distribution is immediate by construction of Algorithm 6. Moreover, note that

$$
\mathbb{E}\left[\omega_{n+1}^1 \mid \mathcal{G}_{n+1}^N \right] = \int w_n(\xi_n^{I_{n+1}^1}, \xi_{n+1}^1, z) \, \mathbf{R}_n(\xi_n^{I_{n+1}^1}, \xi_{n+1}^1, dz) = w_n^\varepsilon(\xi_n^{I_{n+1}^1}, \xi_{n+1}^1),
$$

(B.1)

where w_n and w_n^ε are defined in (**H**2)(ii), and

$$
\mathbb{E}\left[\tau_{n+1}^1 \mid \mathcal{G}_{n+1}^N \right] = \sum_{i=1}^{N} \frac{\omega_n^i \ell_n^\varepsilon(\xi_n^i, \xi_{n+1}^1)}{\sum_{i'=1}^{N} \omega_n^{i'} \ell_n^\varepsilon(\xi_n^{i'}, \xi_{n+1}^1)} \left(\tau_n^i + \tilde{h}_n(\xi_n^i, \xi_{n+1}^1) \right).
$$

Then, since τ_{n+1}^1 and ω_{n+1}^1 are conditionally independent given $\mathcal{G}_{n+1}^N$,

$$
\mathbb{E}\left[\omega_{n+1}^1\{\tau_{n+1}^1 f_{n+1}(\xi_{n+1}^1) + \tilde{f}_{n+1}(\xi_{n+1}^1)\} \mid \mathcal{F}_n^N\right]
$$
$$
= \mathbb{E}\left[\mathbb{E}\left[\omega_{n+1}^1 \mid \mathcal{G}_{n+1}^N\right]\mathbb{E}\left[\tau_{n+1}^1 \mid \mathcal{G}_{n+1}^N\right] f_{n+1}(\xi_{n+1}^1) \mid \mathcal{F}_n^N\right]
$$
$$
+ \mathbb{E}\left[\mathbb{E}\left[\omega_{n+1}^1 \mid \mathcal{G}_{n+1}^N\right]\tilde{f}_{n+1}(\xi_{n+1}^1) \mid \mathcal{F}_n^N\right]
$$
$$
= \mathbb{E}\left[w_n^\varepsilon(\xi_n^{I_{n+1}^1}, \xi_{n+1}^1)\sum_{i=1}^N \frac{\omega_n^i \ell_n^\varepsilon(\xi_n^i, \xi_{n+1}^1)}{\sum_{i'=1}^N \omega_n^{i'} \ell_n^\varepsilon(\xi_n^{i'}, \xi_{n+1}^1)}\left(\tau_n^i + \tilde{h}_n(\xi_n^i, \xi_{n+1}^1)\right) f_{n+1}(\xi_{n+1}^1) \mid \mathcal{F}_n^N\right]
$$
$$
+ \mathbb{E}\left[w_n^\varepsilon(\xi_n^{I_{n+1}^1}, \xi_{n+1}^1)\tilde{f}_{n+1}(\xi_{n+1}^1) \mid \mathcal{F}_n^N\right].
$$

Now, we may complete the proof by noting that, by (2.8),

$$
\mathbb{E}\left[w_n^\varepsilon(\xi_n^{I_{n+1}^1}, \xi_{n+1}^1)\sum_{i=1}^N \frac{\omega_n^i \ell_n^\varepsilon(\xi_n^i, \xi_{n+1}^1)}{\sum_{i'=1}^N \omega_n^{i'} \ell_n^\varepsilon(\xi_n^{i'}, \xi_{n+1}^1)}\left(\tau_n^i + \tilde{h}_n(\xi_n^i, \xi_{n+1}^1)\right) f_{n+1}(\xi_{n+1}^1) \mid \mathcal{F}_n^N\right]
$$
$$
= (\phi_n^N \vartheta_n)^{-1}\sum_{j=1}^N \frac{\omega_n^j}{\Omega_n}\int \ell_n^\varepsilon(\xi_n^j, x)\mu(dx)\sum_{i=1}^N \frac{\omega_n^i \ell_n^\varepsilon(\xi_n^i, x)}{\sum_{i'=1}^N \omega_n^{i'} \ell_n^\varepsilon(\xi_n^{i'}, x)}\left(\tau_n^i + \tilde{h}_n(\xi_n^i, x)\right) f_{n+1}(x)
$$
$$
= (\phi_n^N \vartheta_n)^{-1}\sum_{i=1}^N \frac{\omega_n^i}{\Omega_n}\{\tau_n^i \mathbf{L}_n^\varepsilon f_{n+1}(\xi_n^i) + \mathbf{L}_n^\varepsilon(\tilde{h}_n f_{n+1})(\xi_n^i)\}
$$

and, similarly,

$$
\mathbb{E}\left[w_n^\varepsilon(\xi_n^{I_{n+1}^1}, \xi_{n+1}^1)\tilde{f}_{n+1}(\xi_{n+1}^1) \mid \mathcal{F}_n^N\right] = (\phi_n^N \vartheta_n)^{-1}\sum_{i=1}^N \frac{\omega_n^i}{\Omega_n}\mathbf{L}_n^\varepsilon \tilde{f}_{n+1}(\xi_n^i).
$$

$\square$

Proof of Proposition B.1. We establish (i) by induction over n. Write, using Lemma B.2,

$$
\frac{1}{N}\sum_{i=1}^N \omega_{n+1}^i\{\tau_{n+1}^i f_{n+1}(\xi_{n+1}^i) + \tilde{f}_{n+1}(\xi_{n+1}^i)\} - \frac{\phi_n^\varepsilon \mathbf{L}_n^\varepsilon\{\mathbf{T}_{n+1}^\varepsilon h_{n+1} f_{n+1} + \tilde{f}_{n+1}\}}{\phi_n^\varepsilon \vartheta_n}
$$
$$
= \frac{1}{N}\sum_{i=1}^N \omega_{n+1}^i\{\tau_{n+1}^i f_{n+1}(\xi_{n+1}^i) + \tilde{f}_{n+1}(\xi_{n+1}^i)\}
$$
$$
- \mathbb{E}\left[\omega_{n+1}^1\{\tau_{n+1}^1 f_{n+1}(\xi_{n+1}^1) + \tilde{f}_{n+1}(\xi_{n+1}^1)\} \mid \mathcal{F}_n^N\right]
$$
$$
+ \sum_{j=1}^N \frac{\omega_n^j \vartheta_n(\xi_n^j)}{\sum_{j'=1}^N \omega_n^{j'} \vartheta_n(\xi_n^{j'})}\{\tau_n^j \mathbf{P}_n(w_n^\varepsilon f_{n+1})(\xi_n^j) + \mathbf{P}_n(w_n^\varepsilon \tilde{h}_n f_{n+1} + w_n^\varepsilon \tilde{f}_{n+1})(\xi_n^j)\}
$$
$$
- \frac{\phi_{n-1}^\varepsilon \mathbf{L}_{n-1}^\varepsilon\{\mathbf{T}_n^\varepsilon \mathbf{L}_n^\varepsilon f_{n+1} + \mathbf{L}_n^\varepsilon(\tilde{h}_n f_{n+1} + \tilde{f}_{n+1})\}}{\phi_{n-1}^\varepsilon \mathbf{L}_{n-1}^\varepsilon \vartheta_n}. \quad \text{(B.2)}
$$

Since for all i,

$$|\omega_{n+1}^i\{\tau_{n+1}^i f_{n+1}(\xi_{n+1}^i) + \tilde{f}_{n+1}(\xi_{n+1}^i)\}| \leq \|w_n\|_\infty(\|h_{n+1}\|_\infty\|f_{n+1}\|_\infty + \|\tilde{f}_{n+1}\|_\infty),$$

the Hoeffding inequality for conditional expectations provides constants $(d, \tilde{d}) \in \mathbb{R}_{>0}^2$ such that for all $\epsilon > 0$,

$$\mathbb{P}\left(\left|\frac{1}{N}\sum_{i=1}^N \omega_{n+1}^i\{\tau_{n+1}^i f_{n+1}(\xi_{n+1}^i) + \tilde{f}_{n+1}(\xi_{n+1}^i)\}\right.\right.$$
$$\left.\left. - \mathbb{E}\left[\omega_{n+1}^1\{\tau_{n+1}^1 f_{n+1}(\xi_{n+1}^1) + \tilde{f}_{n+1}(\xi_{n+1}^1)\} \mid \mathcal{F}_n^N\right]\right| \geq \epsilon\right) \leq d\exp\left(-\tilde{d}N\epsilon^2\right).$$

In addition, since ϑ_n, $\vartheta_n \mathbf{P}_n(w_n^\varepsilon f_{n+1})$, and $\vartheta_n \mathbf{P}_n(w_n^\varepsilon \tilde{h}_n f_{n+1} + w_n^\varepsilon \tilde{f}_{n+1})$ all belong to $\mathsf{F}(\mathcal{X}_n)$, there exist, by the induction hypothesis and [13, Lemma 4], $(d', \tilde{d}') \in \mathbb{R}_{>0}^2$ such that for all $\epsilon > 0$,

$$\mathbb{P}\left(\left|\sum_{j=1}^N \frac{\omega_n^j \vartheta_n(\xi_n^j)}{\sum_{j'=1}^N \omega_n^{j'}\vartheta_n(\xi_n^{j'})}\{\tau_n^j \mathbf{P}_n(w_n^\varepsilon f_{n+1})(\xi_n^j) + \mathbf{P}_n(w_n^\varepsilon \tilde{h}_n f_{n+1} + w_n^\varepsilon \tilde{f}_{n+1})(\xi_n^j)\}\right.\right.$$
$$\left.\left. - \frac{\phi_{n-1}^\varepsilon \mathbf{L}_{n-1}^\varepsilon \{\mathbf{T}_n^\varepsilon \mathbf{L}_n^\varepsilon f_{n+1} + \mathbf{L}_n^\varepsilon(\tilde{h}_n f_{n+1} + \tilde{f}_{n+1})\}}{\phi_{n-1}^\varepsilon \mathbf{L}_{n-1}^\varepsilon \vartheta_n}\right| \geq \epsilon\right) \leq d'\exp\left(-\tilde{d}'N\epsilon^2\right).$$

Combining the previous two inequalities completes the proof of (i) at time step $n+1$. The base case $n = 1$ is established using again the decomposition (B.2), the standard Hoeffding inequality for independent and identically distributed variables, [13, Lemma 4], and the fact that $h_0 \equiv 0$.

Final, (ii) follows immediately from (i) and [13, Lemma 4]. $\qquad\square$

Appendix C: Proof of Theorem 4.2

We now turn our focus to the proof of Theorem 4.2, which, just like Theorem 4.1, will be established via the following more general result.

Theorem C.1. *Assume (**H1–2**). Then for all $n \in \mathbb{N}$, $M \in \mathbb{N}_{>0}$, and $(f_n, \tilde{f}_n) \in \mathsf{F}(\mathcal{X}_n)^2$,*

$$\sqrt{N}\left(\sum_{i=1}^N \frac{\omega_n^i}{\Omega_n}\{\tau_n^i f_n(\xi_n^i) + \tilde{f}_n(\xi_n^i)\} - \phi_n^\varepsilon(\mathbf{T}_n^\varepsilon h_n f_n + \tilde{f}_n)\right) \xrightarrow{\mathcal{D}} \sigma_n(h_n, f_n, \tilde{f}_n)Z,$$

where Z is standard normally distributed and

$$\sigma_n^2(h_n, f_n, \tilde{f}_n) := \frac{\chi\{w_{-1}\bar{\mathcal{R}}_{0,n}^\varepsilon(h_n f_n + \tilde{f}_n)\}^2}{(\chi\mathbf{L}_{0,n-1}^\varepsilon \mathbb{1}_{\mathsf{X}^n})^2}$$
$$+ \sigma_n^2\langle(w_\ell^\varepsilon)_{\ell=0}^{n-1}\rangle(h_n, f_n, \tilde{f}_n) + \sigma_n^2\langle(\varsigma_\ell^2)_{\ell=0}^{n-1}\rangle(h_n, f_n, \tilde{f}_n) \quad \text{(C.1)}$$

and

$$\sigma_n^2 \langle (\varphi_\ell)_{\ell=0}^{n-1} \rangle (h_n, f_n, \tilde{f}_n) := \sum_{m=0}^{n-1} \frac{\phi_m^\varepsilon \vartheta_m \phi_m^\varepsilon \mathbf{L}_m^\varepsilon \{\varphi_m[\bar{\mathcal{R}}_{m+1,n}^\varepsilon (h_n f_n + \tilde{f}_n)]^2\}}{(\phi_m^\varepsilon \mathbf{L}_{m,n-1}^\varepsilon \mathbb{1}_{\mathsf{X}_n})^2}$$

$$+ \sum_{m=0}^{n-1} \sum_{\ell=0}^{m} \frac{\phi_m^\varepsilon \vartheta_m \phi_\ell^\varepsilon \mathbf{L}_\ell^\varepsilon \{\mathbf{B}_\ell^\varepsilon (\mathbf{T}_\ell^\varepsilon h_\ell + \tilde{h}_\ell - \mathbf{T}_{\ell+1}^\varepsilon h_{\ell+1})^2 \mathbf{L}_{\ell+1,m}^\varepsilon (\mathbf{B}_m^\varepsilon \varphi_m[\mathbf{L}_{m+1,n-1}^\varepsilon f_n]^2)\}}{M^{m-\ell+1}(\phi_\ell^\varepsilon \mathbf{L}_{\ell,m-1}^\varepsilon \mathbb{1}_{\mathsf{X}_m})(\phi_m^\varepsilon \mathbf{L}_{m,n-1}^\varepsilon \mathbb{1}_{\mathsf{X}_n})^2}$$

$$\text{(C.2)}$$

for any sequence $(\varphi_\ell)_{\ell \in \mathbb{N}}$ *of measurable functions* $\varphi_n : \mathsf{X}_n \times \mathsf{X}_{n+1} \to \mathbb{R}_{\geq 0}$.

As previously, applying Theorem C.1 with $f_n \equiv 1$ and $\tilde{f}_n \equiv 0$ yields immediately Theorem 4.1. Before proving Theorem C.1 we establish some preparatory lemmas, where the first is easily obtained by combining Lemma 2.2(i) (applied to the skew modeled formed by χ and $(\mathbf{L}_n^\varepsilon)_{n \in \mathbb{N}}$) and (2.13).

Lemma C.2. *Assume (**H1**). Then for all* $n \in \mathbb{N}$ *and* $(f_{n+1}, \tilde{f}_{n+1}) \in \mathsf{F}(\mathcal{X}_{n+1})^2$,

$$\phi_n^\varepsilon \mathbf{L}_n^\varepsilon (\mathbf{T}_{n+1}^\varepsilon h_{n+1} f_{n+1} + \tilde{f}_{n+1}) = \phi_n^\varepsilon \{\mathbf{T}_n^\varepsilon \mathbf{L}_n^\varepsilon f_{n+1} + \mathbf{L}_n^\varepsilon (\tilde{h}_n f_{n+1} + \tilde{f}_{n+1})\}.$$

Lemma C.3. *Assume (**H1–2**). Then for all* $n \in \mathbb{N}$, $f \in \mathsf{F}(\mathcal{X}_n)$, *and* $M \in \mathbb{N}_{>0}$,

$$\sum_{i=1}^{N} \frac{\omega_n^i}{\Omega_n} (\tau_n^i)^2 f(\xi_n^i) \xrightarrow{\mathbb{P}} \phi_n^\varepsilon ([\mathbf{T}_n^\varepsilon h_n]^2 f) + \eta_n f,$$

where the measures $(\eta_m)_{m \in \mathbb{N}_{>0}}$ *are defined recursively as*

$$\eta_{m+1} f = M^{-1} \frac{\eta_m \mathbf{L}_m^\varepsilon f + \phi_m^\varepsilon \mathbf{L}_m^\varepsilon \{\mathbf{B}_m^\varepsilon (\mathbf{T}_m^\varepsilon h_m + \tilde{h}_m - \mathbf{T}_{m+1}^\varepsilon h_{m+1})^2 f\}}{\phi_m^\varepsilon \mathbf{L}_m^\varepsilon \mathbb{1}_{\mathsf{X}_{m+1}}}, \quad m \in \mathbb{N},$$

with $\eta_0 \equiv 0$.

Following [29, Lemma 13], the measures $(\eta_m)_{m \in \mathbb{N}_{>0}}$ may be expressed non-recursively as

$$\eta_n f = \sum_{m=0}^{n-1} M^{m-n} \frac{\phi_m^\varepsilon \mathbf{L}_m^\varepsilon \{\mathbf{B}_m^\varepsilon (\mathbf{T}_m^\varepsilon h_m + \tilde{h}_m - \mathbf{T}_{m+1}^\varepsilon h_{m+1})^2 \mathbf{L}_{m+1,n-1}^\varepsilon f\}}{\phi_m^\varepsilon \mathbf{L}_{m,n-1}^\varepsilon \mathbb{1}_{\mathsf{X}_n}}.$$

$$\text{(C.3)}$$

Proof of Lemma C.3. We proceed by induction over n. The base case $n = 0$ is a trivial consequence of the fact that $\mathbf{T}_0 h_0 = 0$ and $\tau_0^i = 0$ for all $i \in [\![1, N]\!]$. Thus, we assume that the result holds true for some $n \in \mathbb{N}$ and write

$$\sum_{i=1}^{N} \frac{\omega_{n+1}^i}{\Omega_{n+1}} (\tau_{n+1}^i)^2 f(\xi_{n+1}^i) = \frac{a_N}{b_N},$$

where

$$a_N := \frac{1}{N} \sum_{i=1}^{N} \omega_{n+1}^i (\tau_{n+1}^i)^2 f(\xi_{n+1}^i), \quad b_N := \frac{1}{N} \sum_{i=1}^{N} \omega_{n+1}^i$$

We first establish the convergence in probability of $(a_N)_{N\in\mathbb{N}_{>0}}$. Using again Hoeffding's inequality for conditional expectations and the fact that the variables $\omega_{n+1}^i(\tau_{n+1}^i)^2 f_{n+1}(\xi_{n+1}^i), i \in [\![1, N]\!]$, are bounded by $\|w_n\|_\infty\|h_{n+1}\|_\infty^2\|f_{n+1}\|_\infty$ and conditionally i.i.d. given $\mathcal{F}_n^N$ for all i, we obtain for all $\epsilon > 0$,

$$\mathbb{P}\left(\left|a_N - \mathbb{E}[a_N \mid \mathcal{F}_n^N]\right| \geq \epsilon\right) \leq \exp\left(-\frac{2N\epsilon^2}{\|w_n\|_\infty\|h_{n+1}\|_\infty^2\|f_{n+1}\|_\infty}\right).$$

It is hence enough to consider the limit in probability of $\mathbb{E}[a_N \mid \mathcal{F}_n^N]$ as N tends to infinity. For this purpose, write, using (B.1),

$$\begin{aligned}
\mathbb{E}[a_N \mid \mathcal{F}_n^N] &= \mathbb{E}\left[\omega_{n+1}^1(\tau_{n+1}^1)^2 f_{n+1}(\xi_{n+1}^1) \mid \mathcal{F}_n^N\right] \\
&= \mathbb{E}\left[w_n^\varepsilon(\xi_n^{I_{n+1}^1}, \xi_{n+1}^1)\mathbb{E}\left[(\tau_{n+1}^1)^2 \mid \mathcal{G}_{n+1}^N\right]f_{n+1}(\xi_{n+1}^1) \mid \mathcal{F}_n^N\right] \\
&= a_N' + a_N'',
\end{aligned}$$

where

$$\begin{aligned}
a_N' &:= M^{-1} \\
&\times \mathbb{E}\left[w_n^\varepsilon(\xi_n^{I_{n+1}^1}, \xi_{n+1}^1)f_{n+1}(\xi_{n+1}^1)\mathbb{E}\left[\left(\tau_n^{J_{n+1}^{(1,1)}} + \tilde{h}_n(\xi_n^{J_{n+1}^{(1,1)}}, \xi_{n+1}^1)\right)^2 \mid \mathcal{G}_{n+1}^N\right] \mid \mathcal{F}_n^N\right], \\
a_N'' &:= (M-1)M^{-1} \\
&\times \mathbb{E}\left[w_n^\varepsilon(\xi_n^{I_{n+1}^1}, \xi_{n+1}^1)f_{n+1}(\xi_{n+1}^1)\mathbb{E}\left[\tau_n^{J_{n+1}^{(1,1)}} + \tilde{h}_n(\xi_n^{J_{n+1}^{(1,1)}}, \xi_{n+1}^1) \mid \mathcal{G}_{n+1}^N\right]^2 \mid \mathcal{F}_n^N\right].
\end{aligned}$$

Here the first term is given by

$$\begin{aligned}
a_N' &= M^{-1}\sum_{i=1}^N \frac{\omega_n^i \vartheta_n(\xi_n^i)}{\sum_{i'=1}^N \omega_n^{i'}\vartheta_n(\xi_n^{i'})}\int w_n^\varepsilon(\xi_n^i, x)f_{n+1}(x) \\
&\qquad\qquad \times \sum_{j=1}^N \frac{\omega_n^j \ell_n^\varepsilon(\xi_n^j, x)}{\sum_{j'=1}^N \omega_n^{j'}\ell_n^\varepsilon(\xi_n^{j'}, x)}\left(\tau_n^j + \tilde{h}_n(\xi_n^j, x)\right)^2 p_n(\xi_n^i, x)\,\mu(dx), \\
&= (M\phi_n^N \vartheta_n)^{-1}\sum_{j=1}^N \frac{\omega_n^j}{\Omega_n}\int f_{n+1}(x)\left(\tau_n^j + \tilde{h}_n(\xi_n^j, x)\right)^2 \mathbf{L}_n^\varepsilon(\xi_n^j, dx). \qquad\qquad \text{(C.4)}
\end{aligned}$$

Thus, by the induction hypothesis and Proposition B.1, a_N' tends in probability to

$$\begin{aligned}
(M\phi_n^\varepsilon \vartheta_n)^{-1}&\left(\eta_n \mathbf{L}_n^\varepsilon f_{n+1} + \phi_n^\varepsilon\{(\mathbf{T}_n^\varepsilon h_n)^2 \mathbf{L}_n^\varepsilon f_{n+1}\}\right. \\
&\left.+ \phi_n^\varepsilon \mathbf{L}_n^\varepsilon(f_{n+1}\tilde{h}_n^2) + 2\phi_n^\varepsilon\{\mathbf{T}_n^\varepsilon h_n \mathbf{L}_n^\varepsilon(f_{n+1}\tilde{h}_n)\}\right) \\
&= (M\phi_n^\varepsilon \vartheta_n)^{-1}\left(\eta_n \mathbf{L}_n^\varepsilon f_{n+1} + \phi_n^\varepsilon \mathbf{L}_n^\varepsilon\{(\mathbf{T}_n^\varepsilon h_n + \tilde{h}_n)^2 f_{n+1}\}\right).
\end{aligned}$$

We turn to the second term. Along the lines of (C.4) we may write

$$a_N'' = (M-1)M^{-1}\frac{\phi_n^N(\mathbf{L}_n^\varepsilon\varphi_N)}{\phi_n^N\vartheta_n}$$

with

$$\varphi_N(x) \coloneqq f_{n+1}(x)\left\{\sum_{i=1}^N \frac{\omega_n^i\ell_n^\varepsilon(\xi_n^i,x)}{\sum_{i'=1}^N \omega_n^{i'}\ell_n^\varepsilon(\xi_n^{i'},x)}\left(\tau_n^i + \tilde{h}_n(\xi_n^i,x)\right)\right\}^2, \quad x \in \mathsf{X}_{n+1}.$$

Now, Proposition B.1 implies pointwise convergence of φ_N in the sense that for all $x \in \mathsf{X}_{n+1}$, $\mathbb{P}$-a.s.,

$$\begin{aligned}
\lim_{N\to\infty}\varphi_N(x) &= f_{n+1}(x)\left(\frac{\phi_n^\varepsilon[\mathbf{T}_n^\varepsilon h_n\ell_n^\varepsilon(\cdot,x) + \tilde{h}_n(\cdot,x)\ell_n^\varepsilon(\cdot,x)]}{\phi_n^\varepsilon[\ell_n^\varepsilon(\cdot,x)]}\right)^2 \\
&= f_{n+1}(x)(\mathbf{B}_n^\varepsilon)^2(\mathbf{T}_n^\varepsilon h_n + \tilde{h}_n)(x) \\
&= f_{n+1}(x)(\mathbf{T}_{n+1}^\varepsilon)^2 h_{n+1}(x).
\end{aligned}$$

Therefore, since $\|\varphi_N\|_\infty \leq \|f_{n+1}\|_\infty\|h_{n+1}\|_\infty^2$ for all N, we may apply Lemma F.1 in order to obtain the limit

$$a_N'' \xrightarrow{\mathbb{P}} (M-1)M^{-1}\frac{\phi_n^\varepsilon\mathbf{L}_n^\varepsilon\{f_{n+1}(\mathbf{T}_{n+1}^\varepsilon)^2 h_{n+1}\}}{\phi_n^\varepsilon\vartheta_n}.$$

Finally, by Proposition B.1(i), $\mathbb{P}$-a.s.,

$$\lim_{N\to\infty} b_N = \frac{\phi_n^\varepsilon\mathbf{L}_n^\varepsilon\mathbb{1}_{\mathsf{X}_{n+1}}}{\phi_n^\varepsilon\vartheta_n}, \tag{C.5}$$

and using (2.4) yields the limit

$$\begin{aligned}
\frac{a_N}{b_N} &\xrightarrow{\mathbb{P}} \phi_{n+1}^\varepsilon\{f_{n+1}(\mathbf{T}_{n+1}^\varepsilon)^2 h_{n+1}\} \\
&+ M^{-1}\frac{\eta_n\mathbf{L}_n^\varepsilon f_{n+1} + \phi_n^\varepsilon\mathbf{L}_n^\varepsilon\{(\mathbf{T}_n^\varepsilon h_n + \tilde{h}_n)^2 f_{n+1}\} - \phi_n^\varepsilon\mathbf{L}_n^\varepsilon\{f_{n+1}(\mathbf{T}_{n+1}^\varepsilon)^2 h_{n+1}\}}{\phi_n^\varepsilon\mathbf{L}_n^\varepsilon\mathbb{1}_{\mathsf{X}_{n+1}}}.
\end{aligned} \tag{C.6}$$

Thus, we may complete the proof by noting, using Lemma 2.2(i), that

$$\begin{aligned}
\phi_n^\varepsilon\mathbf{L}_n^\varepsilon\{(\mathbf{T}_n^\varepsilon h_n + \tilde{h}_n)^2 f_{n+1}\} &- \phi_n^\varepsilon\mathbf{L}_n^\varepsilon\{f_{n+1}(\mathbf{T}_{n+1}^\varepsilon)^2 h_{n+1}\} \\
&= \phi_n^\varepsilon\mathbf{L}_n^\varepsilon\{\mathbf{B}_n^\varepsilon(\mathbf{T}_n^\varepsilon h_n + \tilde{h}_n - \mathbf{T}_{n+1}h_{n+1})^2 f_{n+1}\}.
\end{aligned}$$

$\square$

Proof of Theorem C.1. We proceed by induction over n and assume that the result holds for some $n \in \mathbb{N}$. In addition, we first assume that $\phi_{n+1}^\varepsilon(\mathbf{T}_{n+1}^\varepsilon h_{n+1}f_{n+1}+$

$\tilde{f}_{n+1}) = 0$ (this assumption will be removed in the end of the proof). Write

$$\sqrt{N} \sum_{i=1}^{N} \frac{\omega_{n+1}^i}{\Omega_{n+1}} \{\tau_{n+1}^i f_{n+1}(\xi_{n+1}^i) + \tilde{f}_{n+1}(\xi_{n+1}^i)\} = (\Omega_{n+1}/N)^{-1}(\Delta_N' + \Delta_N'')/N,$$

(C.7)

where

$$\Delta_N' := \sqrt{N} \sum_{i=1}^{N} \mathbb{E}\left[\omega_{n+1}^i \{\tau_{n+1}^i f_{n+1}(\xi_{n+1}^i) + \tilde{f}_{n+1}(\xi_{n+1}^i)\} \mid \mathcal{F}_n^N\right],$$

$$\Delta_N'' := \sqrt{N} \sum_{i=1}^{N} \left(\omega_{n+1}^i \{\tau_{n+1}^i f_{n+1}(\xi_{n+1}^i) + \tilde{f}_{n+1}(\xi_{n+1}^i)\}\right.$$

$$\left. -\mathbb{E}\left[\omega_{n+1}^i \{\tau_{n+1}^i f_{n+1}(\xi_{n+1}^i) + \tilde{f}_{n+1}(\xi_{n+1}^i)\} \mid \mathcal{F}_n^N\right]\right).$$

In the following we establish the weak limit of $(\Delta_N' + \Delta_N'')/N$, from which the weak limit of (C.7) follows by Slutsky's lemma. In order to treat the first term Δ_N'/N of the decomposition (C.7), write, using Lemma B.2,

$$\Delta_N'/N = \sqrt{N}(\phi_n^N \vartheta_n)^{-1} \sum_{i=1}^{N} \frac{\omega_n^i}{\Omega_n} \{\tau_n^i \mathbf{L}_n^\varepsilon f_{n+1}(\xi_n^i) + \mathbf{L}_n^\varepsilon(\tilde{h}_n f_{n+1} + \tilde{f}_{n+1})(\xi_n^i)\}.$$

In the previous expression, $\lim_{N\to\infty} \phi_n^N \vartheta_n = \phi_n^\varepsilon \vartheta_n > 0$, $\mathbb{P}$-a.s., by proposition B.1. In addition, by Lemma 2.2(i),

$$\phi_n^\varepsilon \{\mathbf{T}_n^\varepsilon h_n \mathbf{L}_n^\varepsilon f_{n+1} + \mathbf{L}_n^\varepsilon(\tilde{h}_n f_{n+1} + \tilde{f}_{n+1})\}$$

$$= \iint \phi_n^\varepsilon \mathbf{L}_n^\varepsilon(dx_{n+1}) \mathbf{B}_n^\varepsilon(x_{n+1}, dx_n)$$

$$\times \{\mathbf{T}_n^\varepsilon h_n(x_n) f_{n+1}(x_{n+1}) + \tilde{h}_n(x_n, x_{n+1}) f_{n+1}(x_{n+1}) + \tilde{f}_{n+1}(x_{n+1})\}$$

$$= \phi_n^\varepsilon \mathbf{L}_n^\varepsilon \mathbb{1}_{\mathsf{X}_{n+1}} \times \phi_{n+1}^\varepsilon(\mathbf{T}_{n+1}^\varepsilon h_{n+1} f_{n+1} + \tilde{f}_{n+1}) = 0,$$

where the last equality holds by assumption. Thus, applying the induction hypothesis and Slutsky's lemma yields

$$\lim_{N\to\infty} \mathbb{E}\left[\exp\left(i\, u\Delta_N'/N\right)\right] = \exp\left(-u^2 \frac{\sigma_n^2 \langle \mathbf{L}_n^\varepsilon f_{n+1}; \mathbf{L}_n^\varepsilon(\tilde{h}_n f_{n+1} + \tilde{f}_{n+1})\rangle}{2(\phi_n^\varepsilon \vartheta_n)^2}\right).$$

(C.8)

We turn to the second term of (C.7). By Lemma B.2, $\Delta_N''/N = \sum_{i=1}^{N} \upsilon_N^i$, where

$$\upsilon_N^i := \frac{1}{M\sqrt{N}} \sum_{j=1}^{M} \tilde{\upsilon}_N(I_{n+1}^i, J_{n+1}^{(i,j)}, \xi_{n+1}^i, \zeta_{n+1}^i),$$

with

$$\tilde{v}_N(i,j,x,z) := w_n(\xi_n^i, x, z) \left(\{\tau_n^j + \tilde{h}_n(\xi_n^j, x)\} f_{n+1}(x) + \tilde{f}_{n+1}(x) \right)$$

$$- (\phi_n^N \vartheta_n)^{-1} \sum_{\ell=1}^N \frac{\omega_n^\ell}{\Omega_n} \{\tau_n^\ell \mathbf{L}_n^\varepsilon f_{n+1}(\xi_n^\ell) + \mathbf{L}_n^\varepsilon (\tilde{h}_n f_{n+1} + \tilde{f}_{n+1})(\xi_n^\ell)\},$$

$$(i,j,x,z) \in [\![1,N]\!]^2 \times \mathsf{X}_{n+1} \times \mathsf{Z}_{n+1}.$$

Remark C.4. In the framework of a fully dominated HMM and a PaRIS driven by the standard bootstrap particle filter, which was the setting considered in [28], the function $\tilde{v}_N$ does not, on the contrary to the general case considered here, depend on i. As we will see in the next derivations, this dependence calls for a non-trivial refinement of the proof of [28, Theorem 3].

In the following we establish the weak limit of $\sum_{i=1}^N v_N^i$ using [14, Theorem A.3]. By construction, $\mathbb{E}[v_N^i \mid \mathcal{F}_n^N] = 0$; thus, the condition (31) in the mentioned theorem can be checked by establishing that

$$\sum_{i=1}^N \mathbb{E}[(v_N^i)^2 \mid \mathcal{F}_n^N] = M^{-1} \mathbb{E}\left[\mathbb{E}\left[\tilde{v}_N^2(I_{n+1}^1, J_{n+1}^{(1,1)}, \xi_{n+1}^1, \zeta_{n+1}^1) \mid \mathcal{G}_{n+1}^N \right] \mid \mathcal{F}_n^N \right]$$

$$+ (M-1)M^{-1} \mathbb{E}\left[\mathbb{E}^2\left[\tilde{v}_N(I_{n+1}^1, J_{n+1}^{(1,1)}, \xi_{n+1}^1, \zeta_{n+1}^1) \mid \mathcal{G}_{n+1}^N \right] \mid \mathcal{F}_n^N \right] \quad \text{(C.9)}$$

converges in probability as $N \to \infty$. The first term of (C.9) is given by

$$\mathbb{E}\left[\mathbb{E}\left[\tilde{v}_N^2(I_{n+1}^1, J_{n+1}^{(1,1)}, \xi_{n+1}^1, \zeta_{n+1}^1) \mid \mathcal{G}_{n+1}^N \right] \mid \mathcal{F}_n^N \right]$$

$$= \mathbb{E}\left[\sum_{j=1}^N \frac{\omega_n^j \ell_n^\varepsilon(\xi_n^j, \xi_{n+1}^1)}{\sum_{j'=1}^N \omega_n^{j'} \ell_n^\varepsilon(\xi_n^{j'}, \xi_{n+1}^1)} \int \tilde{v}_N^2(I_{n+1}^1, j, \xi_{n+1}^1, z) \, \mathbf{R}_n(\xi_n^{I_{n+1}^1}, \xi_{n+1}^1, dz) \mid \mathcal{F}_n^N \right],$$

$$= \sum_{i=1}^N \frac{\omega_n^i \vartheta_n(\xi_n^i)}{\sum_{i'=1}^N \omega_n^{i'} \vartheta_n(\xi_n^{i'})}$$

$$\times \int p_n(\xi_n^i, x) \sum_{j=1}^N \frac{\omega_n^j \ell_n^\varepsilon(\xi_n^j, x)}{\sum_{j'=1}^N \omega_n^{j'} \ell_n^\varepsilon(\xi_n^{j'}, x)} \int \tilde{v}_N^2(i,j,x,z) \, \mathbf{R}_n(\xi_n^i, x, dz) \, \mu(dx)$$

$$= (\phi_n^N \vartheta_n)^{-1}(a_N + b_N + c_N)$$

$$- (\phi_n^N \vartheta_n)^{-2} \left(\sum_{\ell=1}^N \frac{\omega_n^\ell}{\Omega_n} \{\tau_n^\ell \mathbf{L}_n^\varepsilon f_{n+1}(\xi_n^\ell) + \mathbf{L}_n^\varepsilon (\tilde{h}_n f_{n+1} + \tilde{f}_{n+1})(\xi_n^\ell)\} \right)^2,$$

$$\text{(C.10)}$$

where

$$a_N := \sum_{j=1}^{N} \frac{\omega_n^j}{\Omega_n} (\tau_n^j)^2 \int \ell_n^\varepsilon(\xi_n^j, x) \varphi_N(x) f_{n+1}^2(x) \, \mu(dx),$$

$$b_N := \sum_{i=1}^{N} \frac{\omega_n^i}{\Omega_n} \int l_n(\xi_n^i, x) \sum_{j=1}^{N} \frac{\omega_n^j \ell_n^\varepsilon(\xi_n^j, x)}{\sum_{j'=1}^{N} \omega_n^{j'} \ell_n^\varepsilon(\xi_n^{j'}, x)} \left(\tilde{h}_n(\xi_n^j, x) f_{n+1}(x) + \tilde{f}_{n+1}(x) \right)^2 \mu(dx),$$

$$c_N := 2 \sum_{j=1}^{N} \frac{\omega_n^j}{\Omega_n} \tau_n^j \int \ell_n^\varepsilon(\xi_n^j, x) \varphi_N(x) f_{n+1}(x) \left(\tilde{h}_n(\xi_n^j, x) f_{n+1}(x) + \tilde{f}_{n+1}(x) \right) \mu(dx),$$

with

$$\varphi_N(x) := \frac{\sum_{i=1}^{N} \omega_n^i l_n(\xi_n^i, x)}{\sum_{j'=1}^{N} \omega_n^{j'} \ell_n^\varepsilon(\xi_n^{j'}, x)}, \quad x \in \mathsf{X}_{n+1},$$

being a random function and

$$l_n(x_n, x_{n+1}) := \int \ell_n \langle z \rangle (x_n, x_{n+1}) w_n(x_n, x_{n+1}, z) \, \mathbf{R}_n(x_n, x_{n+1}, dz),$$

$$(x_n, x_{n+1}) \in \mathsf{X}_n \times \mathsf{X}_{n+1},$$

an unnormalised transition density. Define the deterministic function $\varphi(x) := \phi_n[l_n(\cdot, x)]/\phi_n[\ell_n^\varepsilon(\cdot, x)]$, $x \in \mathsf{X}_{n+1}$; then by Proposition B.1, $\mathbb{P}$-a.s., $\lim_{N \to \infty} \varphi_N(x) = \varphi(x)$ for all x. Thus, since

$$\left| a_N - \sum_{j=1}^{N} \frac{\omega_n^j}{\Omega_n} (\tau_n^j)^2 \int \ell_n^\varepsilon(\xi_n^j, x) f_{n+1}^2(x) \varphi(x) \, \mu(dx) \right|$$

$$\leq \|h_n\|_\infty^2 \sum_{j=1}^{N} \frac{\omega_n^j}{\Omega_n} \int \ell_n^\varepsilon(\xi_n^j, x) f_{n+1}^2(x) |\varphi_N(x) - \varphi(x)| \, \mu(dx) \quad \text{(C.11)}$$

and $f_{n+1}^2(x)|\varphi_N(x) - \varphi(x)| \leq 2\|w_n\|_\infty \|f_{n+1}\|_\infty^2$ for all x, Lemma F.1 implies that the right-hand side of (C.11) tends to zero as $N \to \infty$. Thus, by Lemma C.3,

$$a_N \xrightarrow{\mathbb{P}} \eta_n \mathbf{L}_n^\varepsilon(f_{n+1}^2 \varphi) + \phi_n^\varepsilon \{ (\mathbf{T}_n^\varepsilon)^2 h_n \mathbf{L}_n^\varepsilon(f_{n+1}^2 \varphi) \}.$$

We compute the limit of b_N. For this purpose, define the mapping

$$\psi_N(x) := \sum_{j=1}^{N} \frac{\omega_n^j \ell_n^\varepsilon(\xi_n^j, x)}{\sum_{j'=1}^{N} \omega_n^{j'} \ell_n^\varepsilon(\xi_n^{j'}, x)} \left(\tilde{h}_n(\xi_n^j, x) f_{n+1}(x) + \tilde{f}_{n+1}(x) \right)^2, \quad x \in \mathsf{X}_{n+1}.$$

which tends, for all x, by Proposition B.1, $\mathbb{P}$-a.s. to

$$\psi(x) := \mathbf{B}_n^\varepsilon (\tilde{h}_n f_{n+1} + \tilde{f}_{n+1})^2(x).$$

Moreover, since $\psi_N(x) \leq (\|\tilde{h}_n\|_\infty \|f_{n+1}\|_\infty + \|\tilde{f}_{n+1}\|_\infty)^2$ for all x, Lemma F.1 implies that

$$
\begin{aligned}
b_N \xrightarrow{\mathbb{P}} \int & \phi_n[l_n(\cdot, x)]\psi(x)\,\mu(dx) \\
&= \iint l_n(x_n, x_{n+1})\mathbf{B}_n^\varepsilon(\tilde{h}_n f_{n+1} + \tilde{f}_{n+1})^2(x_{n+1})\,\phi_n^\varepsilon(dx_n)\,\mu(dx_{n+1}) \\
&= \phi_n^\varepsilon \mathbf{L}_n^\varepsilon\{\varphi(\tilde{h}_n f_{n+1} + \tilde{f}_{n+1})^2\},
\end{aligned}
$$

where the last equality is obtained using Lemma 2.2(i).

We turn to the last term, c_N. Since

$$
\left| c_N - 2\sum_{j=1}^N \frac{\omega_n^j}{\Omega_n}\tau_n^j \int \ell_n^\varepsilon(\xi_n^j, x)f_{n+1}(x)\varphi(x)\left(\tilde{h}_n(\xi_n^j, x)f_{n+1}(x) + \tilde{f}_{n+1}(x)\right)\mu(dx) \right|
$$

$$
\leq \|h_n\|_\infty(\|\tilde{h}_n\|_\infty \|f_{n+1}\|_\infty + \|\tilde{f}_{n+1}\|_\infty)
$$

$$
\times \sum_{j=1}^N \frac{\omega_n^j}{\Omega_n}\int \ell_n^\varepsilon(\xi_n^j, x)f_{n+1}(x)|\varphi_N(x) - \varphi(x)|\,\mu(dx),
$$

where the right-hand side tends, by Lemma F.1, to zero in probability as $N \to \infty$, using again Proposition B.1 yields

$$
c_N \xrightarrow{\mathbb{P}} 2\phi_n^\varepsilon\{\mathbf{T}_n^\varepsilon h_n \mathbf{L}_n^\varepsilon[\varphi f_{n+1}(\tilde{h}_n f_{n+1} + \tilde{f}_{n+1})]\}.
$$

Finally, by Proposition B.1, $\mathbb{P}$-a.s.,

$$
\lim_{N \to \infty}\sum_{\ell=1}^N \frac{\omega_n^\ell}{\Omega_n}\{\tau_n^\ell \mathbf{L}_n^\varepsilon f_{n+1}(\xi_n^\ell) + \mathbf{L}_n^\varepsilon(\tilde{h}_n f_{n+1} + \tilde{f}_{n+1})(\xi_n^\ell)\}
$$

$$
= \phi_n^\varepsilon\{\mathbf{T}_n^\varepsilon h_n \mathbf{L}_n^\varepsilon f_{n+1} + \mathbf{L}_n^\varepsilon(\tilde{h}_n f_{n+1} + \tilde{f}_{n+1})\} = 0, \quad \text{(C.12)}
$$

where the last equality follows by Lemma C.2 and assumption, and since also $\lim_{N \to \infty}\phi_n^N \vartheta_n = \phi_n^\varepsilon \vartheta_n$, $\mathbb{P}$-a.s., the second term of (C.10) tends $\mathbb{P}$-a.s. to zero.

To sum up, as $N \to \infty$, the first term of (C.9) satisfies the limit

$$
M^{-1}\mathbb{E}\left[\mathbb{E}\left[\tilde{v}_N^2(I_{n+1}^1, J_{n+1}^{(1,1)}, \xi_{n+1}^1, \zeta_{n+1}^1) \mid \mathcal{G}_{n+1}^N\right] \mid \mathcal{F}_n^N\right]
$$

$$
\xrightarrow{\mathbb{P}} (M\phi_n^\varepsilon \vartheta_n)^{-1}\left(\eta_n \mathbf{L}_n^\varepsilon(f_{n+1}^2\varphi) + \phi_n^\varepsilon \mathbf{L}_n^\varepsilon\{\varphi[(\mathbf{T}_n^\varepsilon h_n + \tilde{h}_n)f_{n+1} + \tilde{f}_{n+1}]^2\}\right).
$$

$$
\text{(C.13)}
$$

We turn to the second term of (C.9) and write

$$\mathbb{E}\left[\mathbb{E}^2\left[\tilde{v}_N(I^1_{n+1}, J^{(1,1)}_{n+1}, \xi^1_{n+1}, \zeta^1_{n+1}) \mid \mathcal{G}^N_{n+1}\right] \mid \mathcal{F}^N_n\right]$$

$$= \mathbb{E}\left[\left(\sum_{j=1}^N \frac{\omega^j_n \ell^\varepsilon_n(\xi^j_n, \xi^1_{n+1})}{\sum_{j'=1}^N \omega^{j'}_n \ell^\varepsilon_n(\xi^{j'}_n, \xi^1_{n+1})}\right.\right.$$

$$\left.\left.\times \int \tilde{v}_N(I^1_{n+1}, j, \xi^1_{n+1}, z) \, \mathbf{R}_n(\xi^{I^1_{n+1}}_n, \xi^1_{n+1}, dz)\right)^2 \mid \mathcal{F}^N_n\right]$$

$$= \sum_{i=1}^N \frac{\omega^i_n \vartheta_n(\xi^i_n)}{\sum_{i'=1}^N \omega^{i'}_n \vartheta_n(\xi^{i'}_n)}$$

$$\times \int \left(\sum_{j=1}^N \frac{\omega^j_n \ell^\varepsilon_n(\xi^j_n, x)}{\sum_{j'=1}^N \omega^{j'}_n \ell^\varepsilon_n(\xi^{j'}_n, x)} \int \tilde{v}_N(i, j, x, z) \, \mathbf{R}_n(\xi^i_n, x, dz)\right)^2 \mathbf{P}_n(\xi^i_n, dx)$$

$$= (\phi^N_n \vartheta_n)^{-1} \sum_{i=1}^N \frac{\omega^i_n}{\Omega_n} \int w^\varepsilon_n(\xi^i_n, x) \gamma^2_N(x) \, \mathbf{L}^\varepsilon_n(\xi^i_n, dx)$$

$$- 2(\phi^N_n \vartheta_n)^{-1}(\phi^N_n \mathbf{L}^\varepsilon_n \gamma_N) \sum_{\ell=1}^N \frac{\omega^\ell_n}{\Omega_n} \{\tau^\ell_n \mathbf{L}^\varepsilon_n f_{n+1}(\xi^\ell_n) + \mathbf{L}^\varepsilon_n(\tilde{h}_n f_{n+1} + \tilde{f}_{n+1})(\xi^\ell_n)\}$$

$$+ (\phi^N_n \vartheta_n)^{-1}\left(\sum_{\ell=1}^N \frac{\omega^\ell_n}{\Omega_n} \{\tau^\ell_n \mathbf{L}^\varepsilon_n f_{n+1}(\xi^\ell_n) + \mathbf{L}^\varepsilon_n(\tilde{h}_n f_{n+1} + \tilde{f}_{n+1})(\xi^\ell_n)\}\right)^2,$$

where

$$\gamma_N(x) := f_{n+1}(x) \sum_{j=1}^N \frac{\omega^j_n \ell^\varepsilon_n(\xi^j_n, x)}{\sum_{j'=1}^N \omega^{j'}_n \ell^\varepsilon_n(\xi^{j'}_n, x)} \left(\tau^j_n + \tilde{h}_n(\xi^j_n, x)\right) + \tilde{f}_{n+1}(x), \quad x \in \mathsf{X}_{n+1}.$$

Since, by Proposition B.1, for all $x \in \mathsf{X}_{n+1}$, $\lim_{N \to \infty} \gamma_N(x) = \gamma(x)$ $\mathbb{P}$-a.s., where

$$\gamma(x) := f_{n+1}(x)\mathbf{B}^\varepsilon_n(\mathbf{T}^\varepsilon_n h_n + \tilde{h}_n)(x) + \tilde{f}_{n+1}(x) = f_{n+1}(x)\mathbf{T}^\varepsilon_{n+1} h_{n+1}(x) + \tilde{f}_{n+1}(x),$$

Lemma F.1 implies that

$$\sum_{i=1}^N \frac{\omega^i_n}{\Omega_n} \int w^\varepsilon_n(\xi^i_n, x) \gamma^2_N(x) \, \mathbf{L}^\varepsilon_n(\xi^i_n, dx)$$

$$\xrightarrow{\mathbb{P}} \iint w^\varepsilon_n(x_n, x_{n+1}) \gamma^2(x_{n+1}) \, \phi^\varepsilon_n(dx_n) \, \mathbf{L}^\varepsilon_n(x_n, dx_{n+1}) \quad \text{(C.14)}$$

and $\phi_n^N \mathbf{L}_n^\varepsilon \gamma_N \xrightarrow{\mathbb{P}} \phi_n^\varepsilon \mathbf{L}_n^\varepsilon \gamma$. Thus, by (C.12–C.14),

$$
\sum_{i=1}^N \mathbb{E}[(v_N^i)^2 \mid \mathcal{F}_n^N] \xrightarrow{\mathbb{P}} \delta_n^2 \langle f_{n+1}, \tilde{f}_{n+1} \rangle(h_n)
$$
$$
:= (M\phi_n^\varepsilon \vartheta_n)^{-1} \left(\eta_n \mathbf{L}_n^\varepsilon(f_{n+1}^2 \varphi) + \phi_n^\varepsilon \mathbf{L}_n^\varepsilon \{ \varphi[(\mathbf{T}_n^\varepsilon h_n + \tilde{h}_n) f_{n+1} + \tilde{f}_{n+1}]^2 \} \right.
$$
$$
\left. + (M-1)\phi_n^\varepsilon \mathbf{L}_n^\varepsilon \{ w_n^\varepsilon (f_{n+1} \mathbf{T}_{n+1}^\varepsilon h_{n+1} + \tilde{f}_{n+1})^2 \} \right),
$$

which verifies condition (i) in [14, Theorem A.3]. Now, with the conditional relative weight variance ς_n^2 defined as in (4.2) it holds that

$$
\varphi(x) - \mathbf{B}_n^\varepsilon w_n^\varepsilon(x) = \frac{\phi_n^\varepsilon[\ell_n^\varepsilon(\cdot, x)\varsigma_n^2(\cdot, x)]}{\phi_n^\varepsilon[\ell_n^\varepsilon(\cdot, x)]} = \mathbf{B}_n^\varepsilon \varsigma_n^2(x).
$$

Then, since, applying Lemma 2.2(i) (twice),

$$
\phi_n^\varepsilon \mathbf{L}_n^\varepsilon \{ \varphi[(\mathbf{T}_n^\varepsilon h_n + \tilde{h}_n) f_{n+1} + \tilde{f}_{n+1}]^2 \} - \phi_n^\varepsilon \mathbf{L}_n^\varepsilon \{ w_n^\varepsilon (f_{n+1} \mathbf{T}_{n+1}^\varepsilon h_{n+1} + \tilde{f}_{n+1})^2 \}
$$
$$
= \phi_n^\varepsilon \mathbf{L}_n^\varepsilon \{ w_n^\varepsilon f_{n+1}^2 \mathbf{B}_n^\varepsilon (\mathbf{T}_n^\varepsilon h_n + \tilde{h}_n - \mathbf{T}_{n+1}^\varepsilon h_{n+1})^2 \}
$$
$$
+ \phi_n^\varepsilon \mathbf{L}_n^\varepsilon \{ \mathbf{B}_n^\varepsilon \varsigma_n^2 [(\mathbf{T}_n^\varepsilon h_n + \tilde{h}_n) f_{n+1} + \tilde{f}_{n+1}]^2 \},
$$

we may express the previous limit as

$$
\delta_n^2 \langle f_{n+1}, \tilde{f}_{n+1} \rangle = (M\phi_n^\varepsilon \vartheta_n)^{-1} \left(\eta_n \mathbf{L}_n^\varepsilon(f_{n+1}^2 \mathbf{B}_n^\varepsilon w_n^\varepsilon) \right.
$$
$$
+ \phi_n^\varepsilon \mathbf{L}_n^\varepsilon \{ w_n^\varepsilon f_{n+1}^2 \mathbf{B}_n^\varepsilon (\mathbf{T}_n^\varepsilon h_n + \tilde{h}_n - \mathbf{T}_{n+1}^\varepsilon h_{n+1})^2 \}
$$
$$
+ M\phi_n^\varepsilon \mathbf{L}_n^\varepsilon \{ w_n^\varepsilon (f_{n+1} \mathbf{T}_{n+1}^\varepsilon h_{n+1} + \tilde{f}_{n+1})^2 \}
$$
$$
\left. + \eta_n \mathbf{L}_n^\varepsilon(f_{n+1}^2 \mathbf{B}_n^\varepsilon \varsigma_n^2) + \phi_n^\varepsilon \mathbf{L}_n^\varepsilon \{ \mathbf{B}_n^\varepsilon \varsigma_n^2 [(\mathbf{T}_n^\varepsilon h_n + \tilde{h}_n) f_{n+1} + \tilde{f}_{n+1}]^2 \} \right).
$$

Condition (ii) in [14, Theorem A.3] is satisfied as an immediate consequence of (**H2**); indeed, letting $d_n := 2\|w_n\|_\infty(\|h_{n+1}\|_\infty\|f_{n+1}\|_\infty + \|\tilde{f}_{n+1}\|_\infty)$ it holds that $|v_N^i| \le d_n/\sqrt{N}$ for all $i \in [\![1, N]\!]$, and consequently, for all $\epsilon > 0$, $\mathbb{P}$-a.s.,

$$
\sum_{i=1}^N \mathbb{E}\left[(v_N^i)^2 \mathbb{1}_{\{|v_N^i| \ge \epsilon\}} \mid \mathcal{F}_n^N \right] \le d_n^2 \mathbb{1}_{\{d_n \ge \epsilon\sqrt{N}\}},
$$

where the right-hand side tends to zero as $N \to \infty$. Thus, [14, Theorem A.3] provides the limit

$$
\mathbb{E}\left[\exp\left(\mathrm{i}\, u \Delta_N''/N \right) \mid \mathcal{F}_n^N \right] \xrightarrow{\mathbb{P}} \exp\left(-u^2 \delta_n^2 \langle f_{n+1}, \tilde{f}_{n+1} \rangle/2 \right).
$$

Now, since $\lim_{N \to \infty} \Omega_{n+1}/N = \phi_n^\varepsilon \mathbf{L}_n^\varepsilon \mathbb{1}_{\mathsf{X}_{n+1}}/\phi_n^\varepsilon \vartheta_n$ $\mathbb{P}$-a.s. by Proposition B.1, we may combine the previous limit with (C.8), [12, Lemma A.5], and Slutsky's

lemma in order to obtain the weak convergence

$$
(\Delta_N' + \Delta_N'')/\Omega_{n+1}
$$
$$
\xrightarrow{\mathcal{D}} N\left(0, \frac{\sigma_n^2(h_n, \mathbf{L}_n^\varepsilon f_{n+1}, \mathbf{L}_n^\varepsilon(\tilde{h}_n f_{n+1} + \tilde{f}_{n+1})) + (\phi_n^\varepsilon \vartheta_n)^2 \delta_n^2 \langle f_{n+1}, \tilde{f}_{n+1}\rangle}{(\phi_n^\varepsilon \mathbf{L}_n^\varepsilon \mathbb{1}_{\mathsf{X}_{n+1}})^2}\right).
$$
(C.15)

In order to treat the general case where $\phi_{n+1}^\varepsilon(\mathbf{T}_{n+1}^\varepsilon h_{n+1} f_{n+1} + \tilde{f}_{n+1})$ is non-zero, define
$$
\bar{f}_{n+1} := \tilde{f}_{n+1} - \phi_{n+1}^\varepsilon(\mathbf{T}_{n+1}^\varepsilon h_{n+1} f_{n+1} + \tilde{f}_{n+1}).
$$
Since the functions f_{n+1} and $\bar{f}_{n+1}$ satisfy $\phi_{n+1}^\varepsilon(\mathbf{T}_{n+1}^\varepsilon h_{n+1} f_{n+1} + \bar{f}_{n+1}) = 0$, (C.15) provides

$$
\sqrt{N}\left(\sum_{i=1}^N \frac{\omega_{n+1}^i}{\Omega_{n+1}}\{\tau_{n+1}^i f_{n+1}(\xi_{n+1}^i) + \tilde{f}_{n+1}(\xi_{n+1}^i)\} - \phi_{n+1}^\varepsilon(\mathbf{T}_{n+1}^\varepsilon h_{n+1} f_{n+1} + \tilde{f}_{n+1})\right)
$$
$$
= \sqrt{N}\left(\sum_{i=1}^N \frac{\omega_{n+1}^i}{\Omega_{n+1}}\{\tau_{n+1}^i f_{n+1}(\xi_{n+1}^i) + \bar{f}_{n+1}(\xi_{n+1}^i)\}\right)
$$
$$
\xrightarrow{\mathcal{D}} N\left(0, \sigma_{n+1}^2(h_{n+1}, f_{n+1}, \tilde{f}_{n+1})\right),
$$

where

$$
\sigma_{n+1}^2(h_{n+1}, f_{n+1}, \tilde{f}_{n+1})
$$
$$
:= \frac{\sigma_n^2(h_n, \mathbf{L}_n^\varepsilon f_{n+1}, \mathbf{L}_n^\varepsilon(\tilde{h}_n f_{n+1} + \bar{f}_{n+1})) + (\phi_n^\varepsilon \vartheta_n)^2 \delta_n^2 \langle f_{n+1}, \bar{f}_{n+1}\rangle}{(\phi_n^\varepsilon \mathbf{L}_n^\varepsilon \mathbb{1}_{\mathsf{X}_{n+1}})^2}
$$
$$
= \frac{\sigma_n^2(h_n, \mathbf{L}_n^\varepsilon f_{n+1}, \mathbf{L}_n^\varepsilon(\tilde{h}_n f_{n+1} + \bar{f}_{n+1}))}{(\phi_n^\varepsilon \mathbf{L}_n^\varepsilon \mathbb{1}_{\mathsf{X}_{n+1}})^2} + \frac{\phi_n^\varepsilon \vartheta_n \eta_n \mathbf{L}_n^\varepsilon(f_{n+1}^2 \mathbf{B}_n^\varepsilon w_n^\varepsilon)}{M(\phi_n^\varepsilon \mathbf{L}_n^\varepsilon \mathbb{1}_{\mathsf{X}_{n+1}})^2}
$$
$$
+ \frac{\phi_n^\varepsilon \vartheta_n \phi_n^\varepsilon \mathbf{L}_n^\varepsilon\{w_n^\varepsilon(f_{n+1}\mathbf{T}_{n+1}^\varepsilon h_{n+1} + \bar{f}_{n+1})^2\}}{(\phi_n^\varepsilon \mathbf{L}_n^\varepsilon \mathbb{1}_{\mathsf{X}_{n+1}})^2}
$$
$$
+ \frac{\phi_n^\varepsilon \vartheta_n \phi_n^\varepsilon \mathbf{L}_n^\varepsilon\{w_n^\varepsilon f_{n+1}^2 \mathbf{B}_n^\varepsilon(\mathbf{T}_n^\varepsilon h_n + \tilde{h}_n - \mathbf{T}_{n+1}^\varepsilon h_{n+1})^2\}}{M(\phi_n^\varepsilon \mathbf{L}_n^\varepsilon \mathbb{1}_{\mathsf{X}_{n+1}})^2}
$$
$$
+ \frac{\phi_n^\varepsilon \vartheta_n \eta_n \mathbf{L}_n^\varepsilon(f_{n+1}^2 \mathbf{B}_n^\varepsilon \varsigma_n^2)}{M(\phi_n^\varepsilon \mathbf{L}_n^\varepsilon \mathbb{1}_{\mathsf{X}_{n+1}})^2} + \frac{\phi_n^\varepsilon \vartheta_n \phi_n^\varepsilon \mathbf{L}_n^\varepsilon\{\mathbf{B}_n^\varepsilon \varsigma_n^2[(\mathbf{T}_n^\varepsilon h_n + \tilde{h}_n)f_{n+1} + \bar{f}_{n+1}]^2\}}{M(\phi_n^\varepsilon \mathbf{L}_n^\varepsilon \mathbb{1}_{\mathsf{X}_{n+1}})^2}.
$$
(C.16)

Since by Lemma 2.2(i),

$$
\phi_n^\varepsilon \mathbf{L}_n^\varepsilon\{w_n^\varepsilon f_{n+1}^2 \mathbf{B}_n^\varepsilon(\mathbf{T}_n^\varepsilon h_n + \tilde{h}_n - \mathbf{T}_{n+1}^\varepsilon h_{n+1})^2\}
$$
$$
= \phi_n^\varepsilon \mathbf{L}_n^\varepsilon\{\mathbf{B}_n^\varepsilon(\mathbf{T}_n^\varepsilon h_n + \tilde{h}_n - \mathbf{T}_{n+1}^\varepsilon h_{n+1})^2 f_{n+1}^2 \mathbf{B}_n^\varepsilon w_n^\varepsilon\},
$$

and

$$\phi_n^\varepsilon \mathbf{L}_n^\varepsilon \{\mathbf{B}_n^\varepsilon \varsigma_n^2 [(\mathbf{T}_n^\varepsilon h_n + \tilde{h}_n) f_{n+1} + \bar{f}_{n+1}]^2\}$$
$$= \phi_n^\varepsilon \mathbf{L}_n^\varepsilon \{\varsigma_n^2 (f_{n+1} \mathbf{T}_{n+1}^\varepsilon h_{n+1} + \bar{f}_{n+1})^2\}$$
$$+ \phi_n^\varepsilon \mathbf{L}_n^\varepsilon \{\mathbf{B}_n^\varepsilon (\mathbf{T}_n^\varepsilon h_n + \tilde{h}_n - \mathbf{T}_{n+1}^\varepsilon h_{n+1})^2 f_{n+1}^2 \mathbf{B}_n^\varepsilon \varsigma_n^2\},$$

inserting the expression of η_n given in (C.3) into (C.16) yields,

$$\sigma_{n+1}^2(h_{n+1}, f_{n+1}, \tilde{f}_{n+1})$$
$$= \frac{\sigma_n^2(h_n, \mathbf{L}_n^\varepsilon f_{n+1}, \mathbf{L}_n^\varepsilon (\tilde{h}_n f_{n+1} + \bar{f}_{n+1}))}{(\phi_n^\varepsilon \mathbf{L}_n^\varepsilon \mathbb{1}_{\mathsf{X}_{n+1}})^2} + \frac{\phi_n^\varepsilon \vartheta_n \phi_n^\varepsilon \mathbf{L}_n^\varepsilon \{w_n^\varepsilon (f_{n+1} \mathbf{T}_{n+1}^\varepsilon h_{n+1} + \bar{f}_{n+1})^2\}}{(\phi_n^\varepsilon \mathbf{L}_n^\varepsilon \mathbb{1}_{\mathsf{X}_{n+1}})^2}$$
$$+ \frac{\phi_n^\varepsilon \vartheta_n}{(\phi_n^\varepsilon \mathbf{L}_n^\varepsilon \mathbb{1}_{\mathsf{X}_{n+1}})^2} \sum_{m=0}^n \frac{\phi_m^\varepsilon \mathbf{L}_m^\varepsilon \{\mathbf{B}_m^\varepsilon (\mathbf{T}_m^\varepsilon h_m + \tilde{h}_m - \mathbf{T}_{m+1}^\varepsilon h_{m+1})^2 \mathbf{L}_{m+1,n}^\varepsilon (f_{n+1}^2 \mathbf{B}_n^\varepsilon w_n^\varepsilon)\}}{M^{n-m+1} \phi_m^\varepsilon \mathbf{L}_{m,n-1}^\varepsilon \mathbb{1}_{\mathsf{X}_n}}$$
$$+ \frac{\phi_n^\varepsilon \vartheta_n \phi_n^\varepsilon \mathbf{L}_n^\varepsilon \{\varsigma_n^2 (f_{n+1} \mathbf{T}_{n+1}^\varepsilon h_{n+1} + \bar{f}_{n+1})^2\}}{(\phi_n^\varepsilon \mathbf{L}_n^\varepsilon \mathbb{1}_{\mathsf{X}_{n+1}})^2}$$
$$+ \frac{\phi_n^\varepsilon \vartheta_n}{(\phi_n^\varepsilon \mathbf{L}_n^\varepsilon \mathbb{1}_{\mathsf{X}_{n+1}})^2} \sum_{m=0}^n \frac{\phi_m^\varepsilon \mathbf{L}_m^\varepsilon \{\mathbf{B}_m^\varepsilon (\mathbf{T}_m^\varepsilon h_m + \tilde{h}_m - \mathbf{T}_{m+1}^\varepsilon h_{m+1})^2 \mathbf{L}_{m+1,n}^\varepsilon (f_{n+1}^2 \mathbf{B}_n^\varepsilon \varsigma_n^2)\}}{M^{n-m+1} \phi_m^\varepsilon \mathbf{L}_{m,n-1}^\varepsilon \mathbb{1}_{\mathsf{X}_n}}.$$
$$\tag{C.17}$$

It remains to establish the non-recursive expression (C.1) of the asymptotic variance. We proceed by induction and assume that (C.1) holds true at time n. Recall the definitions of the kernels $\mathcal{R}_{m,n}^\varepsilon$ and $\bar{\mathcal{R}}_{m,n}^\varepsilon$ in (4.3) and (4.4), respectively. First, note that since by Lemma C.2,

$$\phi_{0:n}^\varepsilon \{h_n \mathbf{L}_n^\varepsilon f_{n+1} + \mathbf{L}_n^\varepsilon (\tilde{h}_n f_{n+1} + \bar{f}_{n+1})\} = 0,$$

it holds for all $m \in [\![0, n-1]\!]$,

$$\bar{\mathcal{R}}_{m,n}^\varepsilon \{h_n \mathbf{L}_n^\varepsilon f_{n+1} + \mathbf{L}_n^\varepsilon (\tilde{h}_n f_{n+1} + \bar{f}_{n+1})\}$$
$$= \mathcal{R}_{m,n}^\varepsilon \{h_n \mathbf{L}_n^\varepsilon f_{n+1} + \mathbf{L}_n^\varepsilon (\tilde{h}_n f_{n+1} + \bar{f}_{n+1})\}$$
$$= \mathcal{R}_{m,n+1}^\varepsilon \{h_{n+1} f_{n+1} + \tilde{f}_{n+1} - \phi_{n+1}^\varepsilon (\mathbf{T}_{n+1}^\varepsilon h_{n+1} f_{n+1} + \tilde{f}_{n+1})\}$$
$$= \bar{\mathcal{R}}_{m,n+1}^\varepsilon (h_{n+1} f_{n+1} + \tilde{f}_{n+1}). \tag{C.18}$$

Further, by definition,

$$f_{n+1} \mathbf{T}_{n+1}^\varepsilon h_{n+1} + \bar{f}_{n+1} = \bar{\mathcal{R}}_{n+1,n+1}^\varepsilon (h_{n+1} f_{n+1} + \tilde{f}_{n+1}). \tag{C.19}$$

Since by (C.18),

$$\frac{\chi(w_{-1} \bar{\mathcal{R}}_{0,n}^\varepsilon \{h_n \mathbf{L}_n^\varepsilon f_{n+1} + \mathbf{L}_n^\varepsilon (\tilde{h}_n f_{n+1} + \bar{f}_{n+1})\})^2}{(\chi \mathbf{L}_{0,n-1}^\varepsilon \mathbb{1}_{\mathsf{X}^n})^2 (\phi_n^\varepsilon \mathbf{L}_n^\varepsilon \mathbb{1}_{\mathsf{X}_{n+1}})^2}$$
$$= \frac{\chi \{w_{-1} \bar{\mathcal{R}}_{0,n+1}^\varepsilon (h_{n+1} f_{n+1} + \tilde{f}_{n+1})\}^2}{(\chi \mathbf{L}_{0,n}^\varepsilon \mathbb{1}_{\mathsf{X}^{n+1}})^2},$$

using the induction hypothesis yields

$$\frac{\sigma_n^2(h_n, \mathbf{L}_n^\varepsilon f_{n+1}, \mathbf{L}_n^\varepsilon(\tilde{h}_n f_{n+1} + \bar{f}_{n+1}))}{(\phi_n^\varepsilon \mathbf{L}_n^\varepsilon \mathbb{1}_{\mathsf{X}_{n+1}})^2} = \frac{\chi\{w_{-1}\bar{\mathcal{R}}_{0,n+1}^\varepsilon(h_{n+1}f_{n+1} + \tilde{f}_{n+1})\}^2}{(\chi\mathbf{L}_{0,n}^\varepsilon \mathbb{1}_{\mathsf{X}^{n+1}})^2}$$

$$+ \frac{\sigma_n^2\langle(w_\ell^\varepsilon)_{\ell=0}^{n-1}\rangle(h_n, \mathbf{L}_n^\varepsilon f_{n+1}, \mathbf{L}_n^\varepsilon(\tilde{h}_n f_{n+1} + \bar{f}_{n+1}))}{(\phi_n^\varepsilon \mathbf{L}_n^\varepsilon \mathbb{1}_{\mathsf{X}_{n+1}})^2}$$

$$+ \frac{\sigma_n^2\langle(\varsigma_\ell^2)_{\ell=0}^{n-1}\rangle(h_n, \mathbf{L}_n^\varepsilon f_{n+1}, \mathbf{L}_n^\varepsilon(\tilde{h}_n f_{n+1} + \bar{f}_{n+1}))}{(\phi_n^\varepsilon \mathbf{L}_n^\varepsilon \mathbb{1}_{\mathsf{X}_{n+1}})^2}.$$

Now, for every $(\varphi_\ell)_{\ell\in\mathbb{N}}$, using (C.2), (C.18), and (C.19),

$$\frac{\sigma_n^2\langle(\varphi_\ell)_{\ell=0}^{n-1}\rangle(h_n, \mathbf{L}_n^\varepsilon f_{n+1}, \mathbf{L}_n^\varepsilon(\tilde{h}_n f_{n+1} + \bar{f}_{n+1}))}{(\phi_n^\varepsilon \mathbf{L}_n^\varepsilon \mathbb{1}_{\mathsf{X}_{n+1}})^2}$$

$$+ \frac{\phi_n^\varepsilon \vartheta_n \phi_n^\varepsilon \mathbf{L}_n^\varepsilon\{\varphi_n(f_{n+1}\mathbf{T}_{n+1}^\varepsilon h_{n+1} + \bar{f}_{n+1})^2\}}{(\phi_n^\varepsilon \mathbf{L}_n^\varepsilon \mathbb{1}_{\mathsf{X}_{n+1}})^2}$$

$$+ \frac{\phi_n^\varepsilon \vartheta_n}{(\phi_n^\varepsilon \mathbf{L}_n^\varepsilon \mathbb{1}_{\mathsf{X}_{n+1}})^2} \sum_{m=0}^n \frac{\phi_m^\varepsilon \mathbf{L}_m^\varepsilon\{\mathbf{B}_m^\varepsilon(\mathbf{T}_m^\varepsilon h_m + \tilde{h}_m - \mathbf{T}_{m+1}^\varepsilon h_{m+1})^2 \mathbf{L}_{m+1,n}^\varepsilon(f_{n+1}^2 \mathbf{B}_n^\varepsilon \varphi_n)\}}{M^{n-m+1}\phi_m^\varepsilon \mathbf{L}_{m,n-1}^\varepsilon \mathbb{1}_{\mathsf{X}_n}}$$

$$= \sum_{m=0}^{n-1} \frac{\phi_m^\varepsilon \vartheta_m \phi_m^\varepsilon \mathbf{L}_m^\varepsilon\{\varphi_m[\bar{\mathcal{R}}_{m+1,n+1}^\varepsilon(h_{n+1}f_{n+1} + \tilde{f}_{n+1})]^2\}}{(\phi_m^\varepsilon \mathbf{L}_{m,n}^\varepsilon \mathbb{1}_{\mathsf{X}_{n+1}})^2}$$

$$+ \sum_{m=0}^{n-1}\sum_{\ell=0}^m \frac{\phi_m^\varepsilon \vartheta_m \phi_\ell^\varepsilon \mathbf{L}_\ell^\varepsilon\{\mathbf{B}_\ell^\varepsilon(\mathbf{T}_\ell^\varepsilon h_\ell + \tilde{h}_\ell - \mathbf{T}_{\ell+1}^\varepsilon h_{\ell+1})^2 \mathbf{L}_{\ell+1,m}^\varepsilon(\mathbf{B}_m^\varepsilon \varphi_m[\mathbf{L}_{m+1,n}^\varepsilon f_{n+1}]^2)\}}{M^{m-\ell+1}(\phi_\ell^\varepsilon \mathbf{L}_{\ell,m-1}^\varepsilon \mathbb{1}_{\mathsf{X}_m})(\phi_m^\varepsilon \mathbf{L}_{m,n}^\varepsilon \mathbb{1}_{\mathsf{X}_{n+1}})^2}$$

$$+ \frac{\phi_n^\varepsilon \vartheta_n \phi_n^\varepsilon \mathbf{L}_n^\varepsilon\{\varphi_n[\bar{\mathcal{R}}_{n+1,n+1}^\varepsilon(h_{n+1}f_{n+1} + \tilde{f}_{n+1})]^2\}}{(\phi_n^\varepsilon \mathbf{L}_n^\varepsilon \mathbb{1}_{\mathsf{X}_{n+1}})^2}$$

$$+ \phi_n^\varepsilon \vartheta_n \sum_{\ell=0}^n \frac{\phi_\ell^\varepsilon \mathbf{L}_\ell^\varepsilon\{\mathbf{B}_\ell^\varepsilon(\mathbf{T}_\ell^\varepsilon h_\ell + \tilde{h}_\ell - \mathbf{T}_{\ell+1}^\varepsilon h_{\ell+1})^2 \mathbf{L}_{\ell+1,n}^\varepsilon(f_{n+1}^2 \mathbf{B}_n^\varepsilon \varphi_n)\}}{M^{n-\ell+1}(\phi_\ell^\varepsilon \mathbf{L}_{\ell,n-1}^\varepsilon \mathbb{1}_{\mathsf{X}_n})(\phi_n^\varepsilon \mathbf{L}_n^\varepsilon \mathbb{1}_{\mathsf{X}_{n+1}})^2},$$

which indeed equals $\sigma_{n+1}^2\langle(\varphi_\ell)_{\ell=0}^n\rangle(h_{n+1}, f_{n+1}, \tilde{f}_{n+1})$. Consequently, (C.17) can be expressed as

$$\sigma_{n+1}^2(h_{n+1}, f_{n+1}, \tilde{f}_{n+1}) = \frac{\chi\{w_{-1}\bar{\mathcal{R}}_{0,n+1}^\varepsilon(h_{n+1}f_{n+1} + \tilde{f}_{n+1})\}^2}{(\chi\mathbf{L}_{0,n}^\varepsilon \mathbb{1}_{\mathsf{X}^{n+1}})^2}$$

$$+ \sigma_{n+1}^2\langle(w_\ell^\varepsilon)_{\ell=0}^n\rangle(h_{n+1}, f_{n+1}, \tilde{f}_{n+1}) + \sigma_{n+1}^2\langle(\varsigma_\ell^2)_{\ell=0}^n\rangle(h_{n+1}, f_{n+1}, \tilde{f}_{n+1}),$$

which was to be established.

Finally, it remains to check the base case $n = 1$. Indeed, since the initial particles are drawn independently from ν (and $h_0 \equiv 0$, $\mathbf{T}_0 h_0 \equiv 0$, and $\eta_0 \equiv 0$), $\sigma_0^2(h_0, f_0, \tilde{f}_0) = \chi\{w_{-1}(\tilde{f}_0 - \chi\tilde{f}_0)\}^2/(\chi\mathbb{1}_{\mathsf{X}_0})^2$ for all $(f_0, \tilde{f}_0) \in \mathsf{F}(\mathcal{X}_0)^2$; thus,

(C.17) provides, since $\chi\mathbf{L}_0^\varepsilon(\tilde{h}_0 f_1 + \bar{f}_1) = 0$ and $\mathbf{L}_0^\varepsilon(\tilde{h}_0 f_1 + \bar{f}_1) = \bar{\boldsymbol{\mathcal{R}}}_{0,1}^\varepsilon(h_1 f_1 + \tilde{f}_1)$,

$$\sigma_1^2(h_1, f_1, \tilde{f}_1) := \frac{\chi\{w_{-1}\bar{\boldsymbol{\mathcal{R}}}_{0,1}^\varepsilon(h_1 f_1 + \tilde{f}_1)\}^2}{(\chi\mathbf{L}_0^\varepsilon\mathbb{1}_{\mathsf{X}_1})^2} + \frac{\phi_0\vartheta_0\phi_0\mathbf{L}_0^\varepsilon\{w_0^\varepsilon(f_1\mathbf{T}_1^\varepsilon h_1 + \bar{f}_1)^2\}}{(\phi_0\mathbf{L}_0^\varepsilon\mathbb{1}_{\mathsf{X}_1})^2}$$

$$+ M^{-1}\frac{\phi_0\vartheta_0\phi_0\mathbf{L}_0^\varepsilon\{\mathbf{B}_0^\varepsilon(\tilde{h}_0 - \mathbf{T}_1^\varepsilon h_1)^2 f_1^2\mathbf{B}_0^\varepsilon w_0^\varepsilon\}}{(\phi_0\mathbf{L}_0^\varepsilon\mathbb{1}_{\mathsf{X}_1})^2} + \frac{\phi_0\vartheta_0\phi_0\mathbf{L}_0^\varepsilon\{\varsigma_0^2(f_1\mathbf{T}_1^\varepsilon h_1 + \bar{f}_1)^2\}}{(\phi_0\mathbf{L}_0^\varepsilon\mathbb{1}_{\mathsf{X}_1})^2}$$

$$+ M^{-1}\frac{\phi_0\vartheta_0\phi_0\mathbf{L}_0^\varepsilon\{\mathbf{B}_0^\varepsilon(\tilde{h}_0 - \mathbf{T}_1^\varepsilon h_1)^2 f_1^2\mathbf{B}_0^\varepsilon\varsigma_0^2\}}{(\phi_0\mathbf{L}_0^\varepsilon\mathbb{1}_{\mathsf{X}_1})^2}, \quad \text{(C.20)}$$

and by noting that $f_1\mathbf{T}_1^\varepsilon h_1 + \bar{f}_1 = \bar{\boldsymbol{\mathcal{R}}}_{1,1}^\varepsilon(h_1 f_1 + \tilde{f}_1)$ we may conclude that the previous quantity coincides with (C.1) for $n = 1$. This completes the proof. $\quad\square$

Appendix D: Proof of Theorem 4.10

We preface the proof of Theorem 4.10 by a few technical lemmas. For each $n \in \mathbb{N}$ and $m \in [\![0, n]\!]$, define the kernel

$$\boldsymbol{\mathcal{L}}_{m,n}^\varepsilon(x_m', dx_{0:n}) := \delta_{x_m'}(dx_m)\,\mathbf{T}_m^\varepsilon(x_m, dx_{0:m-1})\prod_{\ell=m}^{n-1}\mathbf{L}_\ell(x_\ell, dx_{\ell+1}), \quad \text{(D.1)}$$

on $\mathsf{X}_m \times \mathcal{X}^n$ (where $\mathbf{T}_m^\varepsilon$ is defined in (4.1)).

Lemma D.1. *For all $m \in [\![0, n]\!]$ and $h \in \mathsf{F}(\mathcal{X}^n)$,*

$$\phi_{0:m}^\varepsilon\mathbf{L}_{m,n-1}h = \phi_m^\varepsilon\boldsymbol{\mathcal{L}}_{m,n}^\varepsilon h. \quad \text{(D.2)}$$

Proof. The result is obtained easily by applying Lemma 2.2(ii) to the skew model and using definition (D.1). $\quad\square$

The following probability measures play a key role in the following: for $h \in \mathsf{F}(\mathcal{X}_m)$,

$$\varphi_{m,n}^\varepsilon h := \frac{\phi_m^\varepsilon(h \times \boldsymbol{\mathcal{L}}_{m,n}^\varepsilon\mathbb{1}_{\mathsf{X}^n})}{\phi_m^\varepsilon\boldsymbol{\mathcal{L}}_{m,n}^\varepsilon\mathbb{1}_{\mathsf{X}^n}},$$

$$\vec{\varphi}_{m,n}^\varepsilon h := \frac{\phi_m^\varepsilon(h \times \mathbf{L}_m^\varepsilon\boldsymbol{\mathcal{L}}_{m+1,n}^\varepsilon\mathbb{1}_{\mathsf{X}^n})}{\phi_m^\varepsilon\mathbf{L}_m^\varepsilon\boldsymbol{\mathcal{L}}_{m+1,n}^\varepsilon\mathbb{1}_{\mathsf{X}^n}},$$

$$\overleftarrow{\varphi}_{m,n}^\varepsilon h := \frac{\phi_{m-1}^\varepsilon\mathbf{L}_{m-1}(h \times \boldsymbol{\mathcal{L}}_{m,n}^\varepsilon\mathbb{1}_{\mathsf{X}^n})}{\phi_{m-1}^\varepsilon\boldsymbol{\mathcal{L}}_{m-1,n}^\varepsilon\mathbb{1}_{\mathsf{X}^n}}.$$

In addition, for each $k \in [\![0, n-1]\!]$, let

$$\bar{h}_{k|n} : \mathsf{X}^n \ni x_{0:n} \mapsto \tilde{h}_k(x_k, x_{k+1}). \quad \text{(D.3)}$$

denote the extension of $\tilde{h}_k$ to X^n.

Lemma D.2. *Let $n \in \mathbb{N}$ and $m \in [\![1, n]\!]$. Then the following holds true for all $k \in [\![0, n-1]\!]$.*

(i) $\varphi_{m,n}^{\varepsilon} \left(\dfrac{\boldsymbol{\mathcal{L}}_{m,n}^{\varepsilon} \bar{h}_{k|n}}{\boldsymbol{\mathcal{L}}_{m,n}^{\varepsilon} \mathbb{1}_{\mathsf{X}^n}} \right) = \dfrac{\phi_m^{\varepsilon} \boldsymbol{\mathcal{L}}_{m,n}^{\varepsilon} \bar{h}_{k|n}}{\phi_m^{\varepsilon} \boldsymbol{\mathcal{L}}_{m,n}^{\varepsilon} \mathbb{1}_{\mathsf{X}^n}}$ *for $m \in [\![1, n]\!]$,*

(ii) $\vec{\varphi}_{m,n}^{\varepsilon} \left(\dfrac{\boldsymbol{\mathcal{L}}_{m,n}^{\varepsilon} \bar{h}_{k|n}}{\boldsymbol{\mathcal{L}}_{m,n}^{\varepsilon} \mathbb{1}_{\mathsf{X}^n}} \right) = \dfrac{\phi_{m+1}^{\varepsilon} \boldsymbol{\mathcal{L}}_{m+1,n}^{\varepsilon} \bar{h}_{k|n}}{\phi_{m+1}^{\varepsilon} \boldsymbol{\mathcal{L}}_{m+1,n}^{\varepsilon} \mathbb{1}_{\mathsf{X}^n}}$ *for $m \in [\![k+1, n]\!]$, and*

(iii) $\overleftarrow{\varphi}_{m,n}^{\varepsilon} \left(\dfrac{\boldsymbol{\mathcal{L}}_{m,n}^{\varepsilon} \bar{h}_{k|n}}{\boldsymbol{\mathcal{L}}_{m,n}^{\varepsilon} \mathbb{1}_{\mathsf{X}^n}} \right) = \dfrac{\phi_{m-1}^{\varepsilon} \boldsymbol{\mathcal{L}}_{m-1,n}^{\varepsilon} \bar{h}_{k|n}}{\phi_{m-1}^{\varepsilon} \boldsymbol{\mathcal{L}}_{m-1,n}^{\varepsilon} \mathbb{1}_{\mathsf{X}^n}}$ *for all $m \in [\![1, k]\!]$.*

Proof. The identity (i) follows straightforwardly by the definition of $\varphi_{m,n}^{\varepsilon}$.

We hence turn to (ii), which is established by first noting that for all $m \in [\![k+1, n]\!]$,

$$\frac{\boldsymbol{\mathcal{L}}_{m,n}^{\varepsilon} \bar{h}_{k|n}}{\boldsymbol{\mathcal{L}}_{m,n}^{\varepsilon} \mathbb{1}_{\mathsf{X}^n}} = \mathbf{B}_{m-1}^{\varepsilon} \cdots \mathbf{B}_{k+1}^{\varepsilon} (\mathbf{B}_k^{\varepsilon} \tilde{h}_k).$$

Now, since, by applying Lemma 2.2(i) to the skew model,

$$
\begin{aligned}
\vec{\varphi}_{m,n}^{\varepsilon} h &= \iint \frac{\phi_m^{\varepsilon}(dx_m)\, \mathbf{L}_m^{\varepsilon}(x_m, dx_{m+1})\, h(x_m) \boldsymbol{\mathcal{L}}_{m+1,n}^{\varepsilon} \mathbb{1}_{\mathsf{X}^n}(x_{m+1})}{\phi_m^{\varepsilon} \mathbf{L}_m^{\varepsilon} \boldsymbol{\mathcal{L}}_{m+1,n}^{\varepsilon} \mathbb{1}_{\mathsf{X}^n}} \\
&= \iint \frac{\phi_m^{\varepsilon} \mathbf{L}_m^{\varepsilon}(dx_{m+1})\, \mathbf{B}_m^{\varepsilon}(x_{m+1}, dx_m)\, h(x_m) \boldsymbol{\mathcal{L}}_{m+1,n}^{\varepsilon} \mathbb{1}_{\mathsf{X}^n}(x_{m+1})}{\phi_m^{\varepsilon} \mathbf{L}_m^{\varepsilon} \boldsymbol{\mathcal{L}}_{m+1,n}^{\varepsilon} \mathbb{1}_{\mathsf{X}^n}} \\
&= \frac{\phi_{m+1}^{\varepsilon} (\mathbf{B}_m^{\varepsilon} h \times \boldsymbol{\mathcal{L}}_{m+1,n}^{\varepsilon} \mathbb{1}_{\mathsf{X}^n})}{\phi_{m+1}^{\varepsilon} \boldsymbol{\mathcal{L}}_{m+1,n}^{\varepsilon} \mathbb{1}_{\mathsf{X}^n}},
\end{aligned}
$$

we may establish the identity by proceeding like

$$
\begin{aligned}
&\vec{\varphi}_{m,n}^{\varepsilon} \left(\frac{\boldsymbol{\mathcal{L}}_{m,n}^{\varepsilon} \bar{h}_{k|n}}{\boldsymbol{\mathcal{L}}_{m,n}^{\varepsilon} \mathbb{1}_{\mathsf{X}^n}} \right) \\
&= \iint \frac{\phi_{m+1}^{\varepsilon}(dx_{m+1})\, \mathbf{B}_m^{\varepsilon}(x_{m+1}, dx_m)\, \mathbf{B}_{m-1}^{\varepsilon} \cdots \mathbf{B}_{k+1}^{\varepsilon} (\mathbf{B}_k^{\varepsilon} \tilde{h}_k)(x_m) \boldsymbol{\mathcal{L}}_{m+1,n}^{\varepsilon} \mathbb{1}_{\mathsf{X}^n}(x_{m+1})}{\phi_{m+1}^{\varepsilon} \boldsymbol{\mathcal{L}}_{m+1,n}^{\varepsilon} \mathbb{1}_{\mathsf{X}^n}} \\
&= \frac{\phi_{m+1}^{\varepsilon} \boldsymbol{\mathcal{L}}_{m+1,n}^{\varepsilon} \bar{h}_{k|n}}{\phi_{m+1}^{\varepsilon} \boldsymbol{\mathcal{L}}_{m+1,n}^{\varepsilon} \mathbb{1}_{\mathsf{X}^n}}.
\end{aligned}
$$

Finally, to check (iii), note that for all $m \in [\![1, k]\!]$,

$$\mathbf{L}_{m-1} \boldsymbol{\mathcal{L}}_{m,n}^{\varepsilon} \bar{h}_{k|n} = \boldsymbol{\mathcal{L}}_{m-1,n}^{\varepsilon} \bar{h}_{k|n}.$$

Thus, in this case

$$\overleftarrow{\varphi}_{m,n}^{\varepsilon} \left(\frac{\boldsymbol{\mathcal{L}}_{m,n}^{\varepsilon} \bar{h}_{k|n}}{\boldsymbol{\mathcal{L}}_{m,n}^{\varepsilon} \mathbb{1}_{\mathsf{X}^n}} \right) = \frac{\phi_{m-1}^{\varepsilon} \mathbf{L}_{m-1} \boldsymbol{\mathcal{L}}_{m,n}^{\varepsilon} \bar{h}_{k|n}}{\phi_{m-1}^{\varepsilon} \boldsymbol{\mathcal{L}}_{m-1,n}^{\varepsilon} \mathbb{1}_{\mathsf{X}^n}} = \frac{\phi_{m-1}^{\varepsilon} \boldsymbol{\mathcal{L}}_{m-1,n}^{\varepsilon} \bar{h}_{k|n}}{\phi_{m-1}^{\varepsilon} \boldsymbol{\mathcal{L}}_{m-1,n}^{\varepsilon} \mathbb{1}_{\mathsf{X}^n}}.$$

$\square$

Lemma D.3. *Assume (**H3**). Then for all $n \in \mathbb{N}$, $m \in [\![0, n]\!]$, $k \in [\![0, n-1]\!]$, and $(\lambda, \lambda') \in \mathsf{M}_1(\mathcal{X}_m)^2$,*

$$\left| \frac{\lambda \boldsymbol{\mathcal{L}}_{m,n}^{\varepsilon} \bar{h}_{k|n}}{\lambda \boldsymbol{\mathcal{L}}_{m,n}^{\varepsilon} \mathbb{1}_{\mathsf{X}^n}} - \frac{\lambda' \boldsymbol{\mathcal{L}}_{m,n}^{\varepsilon} \bar{h}_{k|n}}{\lambda' \boldsymbol{\mathcal{L}}_{m,n}^{\varepsilon} \mathbb{1}_{\mathsf{X}^n}} \right| \leq \|\tilde{h}_k\|_{\infty} \rho^{|k-m|-1}.$$

Proof. First, assume that $m \leq k$; then note that for all $x_m \in \mathsf{X}_m$,

$$\frac{\boldsymbol{\mathcal{L}}_{m,n}^{\varepsilon} \bar{h}_{k|n}(x_m)}{\boldsymbol{\mathcal{L}}_{m,n}^{\varepsilon} \mathbb{1}_{\mathsf{X}^n}(x_m)} - \frac{\boldsymbol{\mathcal{L}}_{m,n}^{\varepsilon} \bar{h}_{k|n}(x_m)}{\boldsymbol{\mathcal{L}}_{m,n}^{\varepsilon} \mathbb{1}_{\mathsf{X}^n}(x_m)} = \mathbf{F}_{m|n} \cdots \mathbf{F}_{k-1|n}(\mathbf{F}_{k|n} \tilde{h}_k)(x_m)$$
$$- \mathbf{F}_{m|n} \cdots \mathbf{F}_{k-1|n}(\mathbf{F}_{k|n} \tilde{h}_k)(x_m), \quad \text{(D.4)}$$

where we have introduced the *forward kernels*

$$\mathbf{F}_{m|n} h(x_m) := \frac{\mathbf{L}_m(h \times \boldsymbol{\mathcal{L}}_{m+1,n}^{\varepsilon} \mathbb{1}_{\mathsf{X}^n})(x_m)}{\boldsymbol{\mathcal{L}}_{m,n}^{\varepsilon} \mathbb{1}_{\mathsf{X}^n}(x_m)}, \quad x_m \in \mathsf{X}_m, \quad h \in \mathsf{F}(\mathcal{X}_{m+1}).$$

Under (**H3**), each forward kernel satisfies a global Doeblin condition in the form of the uniform lower bound

$$\mathbf{F}_{m|n} h(x_m) \geq \frac{\sigma_-}{\sigma_+} \mu_{m,n} h,$$

where we have defined the probability measure

$$\mu_{m,n} h := \frac{\mu_{m+1}(h \times \boldsymbol{\mathcal{L}}_{m+1,n}^{\varepsilon} \mathbb{1}_{\mathsf{X}^n})}{\mu_{m+1} \boldsymbol{\mathcal{L}}_{m+1,n}^{\varepsilon} \mathbb{1}_{\mathsf{X}^n}}, \quad h \in \mathsf{F}(\mathcal{X}_{m+1})$$

(where μ_{m+1} is the reference measure introduced in Section 2.1). Thus, by standard results for uniformly minorised Markov chains (see, *e.g.*, [7, Lemma 4.3.13]), the Dobrushin coefficient of each $\mathbf{F}_{m|n}$ is bounded by $\rho = 1 - \sigma_-/\sigma_+$. Thus, (D.4) implies that

$$\left| \frac{\lambda \boldsymbol{\mathcal{L}}_{m,n}^{\varepsilon} \bar{h}_{k|n}}{\lambda \boldsymbol{\mathcal{L}}_{m,n}^{\varepsilon} \mathbb{1}_{\mathsf{X}^n}} - \frac{\lambda' \boldsymbol{\mathcal{L}}_{m,n}^{\varepsilon} \bar{h}_{k|n}}{\lambda' \boldsymbol{\mathcal{L}}_{m,n}^{\varepsilon} \mathbb{1}_{\mathsf{X}^n}} \right| = \left| (\lambda_{m,n} - \lambda'_{m,n}) \mathbf{F}_{m|n} \cdots \mathbf{F}_{k-1|n}(\mathbf{F}_{k|n} \tilde{h}_k) \right|$$
$$\leq \rho^{k-m} \|\mathbf{F}_{k|n} \tilde{h}_k\|_{\infty} \leq \rho^{k-m} \|\tilde{h}_k\|_{\infty},$$

where for $h \in \mathsf{F}(\mathcal{X}_m)$,

$$\lambda_{m,n} h := \frac{\lambda(h \times \boldsymbol{\mathcal{L}}_{m,n}^{\varepsilon} \mathbb{1}_{\mathsf{X}^n})}{\lambda \boldsymbol{\mathcal{L}}_{m,n}^{\varepsilon} \mathbb{1}_{\mathsf{X}^n}}, \quad \lambda'_{m,n} h := \frac{\lambda'(h \times \boldsymbol{\mathcal{L}}_{m,n}^{\varepsilon} \mathbb{1}_{\mathsf{X}^n})}{\lambda' \boldsymbol{\mathcal{L}}_{m,n}^{\varepsilon} \mathbb{1}_{\mathsf{X}^n}}.$$

Now, assume that $m > k$; then note that for all $x_m \in \mathsf{X}_m$,

$$\frac{\boldsymbol{\mathcal{L}}_{m,n}^{\varepsilon} \bar{h}_{k|n}(x_m)}{\boldsymbol{\mathcal{L}}_{m,n}^{\varepsilon} \mathbb{1}_{\mathsf{X}^n}(x_m)} - \frac{\boldsymbol{\mathcal{L}}_{m,n}^{\varepsilon} \bar{h}_{k|n}(x_m)}{\boldsymbol{\mathcal{L}}_{m,n}^{\varepsilon} \mathbb{1}_{\mathsf{X}^n}(x_m)}$$
$$= \mathbf{B}_{m-1}^{\varepsilon} \cdots \mathbf{B}_{k+1}^{\varepsilon}(\mathbf{B}_k^{\varepsilon} \tilde{h}_k)(x_m) - \mathbf{B}_{m-1}^{\varepsilon} \cdots \mathbf{B}_{k+1}^{\varepsilon}(\mathbf{B}_k^{\varepsilon} \tilde{h}_k)(x_m). \quad \text{(D.5)}$$

Under (**H3**), also each backward kernel satisfies a Doeblin condition, namely

$$\mathbf{B}_m^\varepsilon h(x_{m+1}) \geq \frac{\sigma_-}{\sigma_+}\phi_m^\varepsilon h,$$

with the marginal ϕ_m^ε playing the role as minorising measure. Thus, the backward kernel Dobrushin coefficients are bounded by the same constant ρ, implying, via (D.5), that

$$\left|\frac{\lambda\boldsymbol{\mathcal{L}}_{m,n}^\varepsilon \bar{h}_{k|n}}{\lambda\boldsymbol{\mathcal{L}}_{m,n}^\varepsilon \mathbb{1}_{\mathsf{X}^n}} - \frac{\lambda'\boldsymbol{\mathcal{L}}_{m,n}^\varepsilon \bar{h}_{k|n}}{\lambda'\boldsymbol{\mathcal{L}}_{m,n}^\varepsilon \mathbb{1}_{\mathsf{X}^n}}\right| = \left|(\lambda_{m,n} - \lambda'_{m,n})\mathbf{B}_{m-1}^\varepsilon \cdots \mathbf{B}_{k+1}^\varepsilon (\mathbf{B}_k^\varepsilon \tilde{h}_k)\right|$$
$$\leq \rho^{m-k-1}\|\tilde{h}_k\|_\infty.$$

This completes the proof. $\qquad\qquad\square$

Lemma D.4. *Assume (**H3**) and (**H4**). Then the following holds.*

(i) *For all $n \in \mathbb{N}$, $m \in [\![0,n]\!]$, $\varepsilon \in \mathcal{E}$, and $h \in \mathsf{F}(\mathcal{X}^n)$,*

$$\left|\varphi_{m,n}^\varepsilon h - \overleftarrow{\varphi}_{m,n}^\varepsilon h\right| \vee \left|\vec{\varphi}_{m,n}^\varepsilon h - \varphi_{m,n}^\varepsilon h\right| \leq 2c\varepsilon\frac{\sigma_+}{\sigma_-^2}\|h\|_\infty.$$

(ii) *For all $n \in \mathbb{N}$, $m \in [\![0,n-1]\!]$, and $\varepsilon \in \mathcal{E}$,*

$$\left|\frac{\phi_{m+1}^\varepsilon\boldsymbol{\mathcal{L}}_{m+1,n}^\varepsilon \bar{h}_{m|n}}{\phi_{m+1}^\varepsilon\boldsymbol{\mathcal{L}}_{m+1,n}^\varepsilon \mathbb{1}_{\mathsf{X}^n}} - \frac{\phi_m^\varepsilon\boldsymbol{\mathcal{L}}_{m,n}^\varepsilon \bar{h}_{m|n}}{\phi_m^\varepsilon\boldsymbol{\mathcal{L}}_{m,n}^\varepsilon \mathbb{1}_{\mathsf{X}^n}}\right| \leq 2c\varepsilon\frac{\sigma_+}{\sigma_-^2}\|h_m\|_\infty.$$

*In both cases, c is the constant in (**H4**).*

Proof. We start with (i). Combining the decomposition

$$\varphi_{m,n}^\varepsilon h - \overleftarrow{\varphi}_{m,n}^\varepsilon h = \varphi_{m,n}^\varepsilon h \left(\frac{\phi_{m-1}^\varepsilon\mathbf{L}_{m-1}\boldsymbol{\mathcal{L}}_{m,n}^\varepsilon \mathbb{1}_{\mathsf{X}^n} - \phi_{m-1}^\varepsilon\mathbf{L}_{m-1}^\varepsilon\boldsymbol{\mathcal{L}}_{m,n}^\varepsilon \mathbb{1}_{\mathsf{X}^n}}{\phi_{m-1}^\varepsilon\boldsymbol{\mathcal{L}}_{m-1,n}^\varepsilon \mathbb{1}_{\mathsf{X}^n}}\right)$$
$$+ \frac{\phi_{m-1}^\varepsilon\mathbf{L}_{m-1}^\varepsilon(h \times \boldsymbol{\mathcal{L}}_{m,n}^\varepsilon \mathbb{1}_{\mathsf{X}^n}) - \phi_{m-1}^\varepsilon\mathbf{L}_{m-1}(h \times \boldsymbol{\mathcal{L}}_{m,n}^\varepsilon \mathbb{1}_{\mathsf{X}^n})}{\phi_{m-1}^\varepsilon\boldsymbol{\mathcal{L}}_{m-1,n}^\varepsilon \mathbb{1}_{\mathsf{X}^n}}$$

with (**H4**) provides the bound

$$\left|\varphi_{m,n}^\varepsilon h - \overleftarrow{\varphi}_{m,n}^\varepsilon h\right| \leq 2c\varepsilon\frac{\|\boldsymbol{\mathcal{L}}_{m,n}^\varepsilon \mathbb{1}_{\mathsf{X}^n}\|_\infty\|h\|_\infty}{\phi_{m-1}^\varepsilon\boldsymbol{\mathcal{L}}_{m-1,n}^\varepsilon \mathbb{1}_{\mathsf{X}^n}}$$

(where c is the constant in (**H4**)). Now, since for all $x_m \in \mathsf{X}_m$,

$$\boldsymbol{\mathcal{L}}_{m,n}^\varepsilon \mathbb{1}_{\mathsf{X}^n}(x_m) = \int \ell_m(x_m, x_{m+1})\boldsymbol{\mathcal{L}}_{m+1,n}^\varepsilon \mathbb{1}_{\mathsf{X}^n}(x_{m+1})\,\mu_{m+1}(dx_{m+1})$$
$$\leq \sigma_+\mu_{m+1}\boldsymbol{\mathcal{L}}_{m+1,n}^\varepsilon \mathbb{1}_{\mathsf{X}^n}$$

and

$$\phi^\varepsilon_{m-1}\boldsymbol{\mathcal{L}}^\varepsilon_{m-1,n}\mathbb{1}_{\mathsf{X}^n} = \iint \phi^\varepsilon_{m-1}(dx_{m-1})\,\ell_{m-1}(x_{m-1},x_m)\ell_m(x_m,x_{m+1})$$
$$\times \boldsymbol{\mathcal{L}}^\varepsilon_{m+1,n}\mathbb{1}_{\mathsf{X}^n}(x_{m+1})\,\mu_m \otimes \mu_{m+1}(dx_{m:m+1})$$
$$\geq \sigma^2_-\mu_{m+1}\boldsymbol{\mathcal{L}}^\varepsilon_{m+1,n}\mathbb{1}_{\mathsf{X}^n},$$

implying the bound

$$\frac{\|\boldsymbol{\mathcal{L}}^\varepsilon_{m,n}\mathbb{1}_{\mathsf{X}^n}\|_\infty}{\phi^\varepsilon_{m-1}\mathbf{L}_{m-1}\boldsymbol{\mathcal{L}}^\varepsilon_{m,n}\mathbb{1}_{\mathsf{X}^n}} \leq \frac{\sigma_+}{\sigma^2_-}, \tag{D.6}$$

we may conclude that $|\varphi^\varepsilon_{m,n}h - \overleftarrow{\varphi}^\varepsilon_{m,n}h| \leq 2c\varepsilon\sigma_+\|h\|_\infty/\sigma^2_-$. Along similar lines, the second difference can be bounded by the same quantity by writing

$$\vec{\varphi}^\varepsilon_{m,n}h - \varphi^\varepsilon_{m,n}h = \vec{\varphi}^\varepsilon_{m,n}h \left(\frac{\phi^\varepsilon_m\mathbf{L}_m\boldsymbol{\mathcal{L}}^\varepsilon_{m+1,n}\mathbb{1}_{\mathsf{X}^n} - \phi^\varepsilon_m\mathbf{L}^\varepsilon_m\boldsymbol{\mathcal{L}}^\varepsilon_{m+1,n}\mathbb{1}_{\mathsf{X}_n}}{\phi^\varepsilon_m\mathbf{L}_m\boldsymbol{\mathcal{L}}^\varepsilon_{m+1,n}\mathbb{1}_{\mathsf{X}^n}} \right)$$
$$+ \frac{\phi^\varepsilon_m(h \times \mathbf{L}^\varepsilon_m\boldsymbol{\mathcal{L}}^\varepsilon_{m+1,n}\mathbb{1}_{\mathsf{X}^n}) - \phi^\varepsilon_m(h \times \mathbf{L}_m\boldsymbol{\mathcal{L}}^\varepsilon_{m+1,n}\mathbb{1}_{\mathsf{X}_n})}{\phi^\varepsilon_m\mathbf{L}_m\boldsymbol{\mathcal{L}}^\varepsilon_{m+1,n}\mathbb{1}_{\mathsf{X}^n}}$$

and reapplying (**H4**) and (D.6).

We turn to (ii). By definition (D.1),

$$\boldsymbol{\mathcal{L}}^\varepsilon_{m+1,n}\bar{h}_{m|n}(x_{m+1}) = \int \tilde{h}_m(x_m,x_{m+1})\,\mathbf{B}^\varepsilon_m(x_{m+1},dx_m)\,\boldsymbol{\mathcal{L}}^\varepsilon_{m+1,n}\mathbb{1}_{\mathsf{X}^n}(x_{m+1});$$

thus, by applying Lemma 2.2(i) to the skew model,

$$\frac{\phi^\varepsilon_{m+1}\boldsymbol{\mathcal{L}}^\varepsilon_{m+1,n}\bar{h}_{m|n}}{\phi^\varepsilon_{m+1}\boldsymbol{\mathcal{L}}^\varepsilon_{m+1,n}\mathbb{1}_{\mathsf{X}^n}} \tag{D.7}$$
$$= \iint \frac{\phi^\varepsilon_m\mathbf{L}^\varepsilon_m(dx_{m+1})\,\tilde{h}_m(x_m,x_{m+1})\,\mathbf{B}^\varepsilon_m(x_{m+1},dx_m)\,\boldsymbol{\mathcal{L}}^\varepsilon_{m+1,n}\mathbb{1}_{\mathsf{X}^n}(x_{m+1})}{\phi^\varepsilon_m\mathbf{L}^\varepsilon_m\boldsymbol{\mathcal{L}}^\varepsilon_{m+1,n}\mathbb{1}_{\mathsf{X}^n}}$$
$$= \iint \frac{\phi^\varepsilon_m(dx_m)\,\mathbf{L}^\varepsilon_m(x_m,dx_{m+1})\,\tilde{h}_m(x_m,x_{m+1})\boldsymbol{\mathcal{L}}^\varepsilon_{m+1,n}\mathbb{1}_{\mathsf{X}^n}(x_{m+1})}{\phi^\varepsilon_m\mathbf{L}^\varepsilon_m\boldsymbol{\mathcal{L}}^\varepsilon_{m+1,n}\mathbb{1}_{\mathsf{X}^n}}.$$

We may thus decompose the quantity under consideration as

$$\frac{\phi^\varepsilon_{m+1}\boldsymbol{\mathcal{L}}^\varepsilon_{m+1,n}\bar{h}_{m|n}}{\phi^\varepsilon_{m+1}\boldsymbol{\mathcal{L}}^\varepsilon_{m+1,n}\mathbb{1}_{\mathsf{X}^n}} - \frac{\phi^\varepsilon_m\boldsymbol{\mathcal{L}}^\varepsilon_{m,n}\bar{h}_{m|n}}{\phi^\varepsilon_m\boldsymbol{\mathcal{L}}^\varepsilon_{m,n}\mathbb{1}_{\mathsf{X}^n}}$$
$$= \frac{\phi^\varepsilon_{m+1}\boldsymbol{\mathcal{L}}^\varepsilon_{m+1,n}\bar{h}_{m|n}}{\phi^\varepsilon_{m+1}\boldsymbol{\mathcal{L}}^\varepsilon_{m+1,n}\mathbb{1}_{\mathsf{X}^n}} \left(\frac{\phi^\varepsilon_m\mathbf{L}_m\boldsymbol{\mathcal{L}}^\varepsilon_{m+1,n}\mathbb{1}_{\mathsf{X}^n} - \phi^\varepsilon_m\mathbf{L}^\varepsilon_m\boldsymbol{\mathcal{L}}^\varepsilon_{m+1,n}\mathbb{1}_{\mathsf{X}^n}}{\phi^\varepsilon_m\boldsymbol{\mathcal{L}}^\varepsilon_{m,n}\mathbb{1}_{\mathsf{X}^n}} \right)$$
$$+ \int \frac{\phi^\varepsilon_m(dx_m)\{\mathbf{L}^\varepsilon_m(\tilde{h}_m\boldsymbol{\mathcal{L}}^\varepsilon_{m+1,n}\mathbb{1}_{\mathsf{X}^n})(x_m) - \mathbf{L}_m(\tilde{h}_m\boldsymbol{\mathcal{L}}^\varepsilon_{m+1,n}\mathbb{1}_{\mathsf{X}^n})(x_m)\}}{\phi^\varepsilon_m\boldsymbol{\mathcal{L}}^\varepsilon_{m,n}\mathbb{1}_{\mathsf{X}^n}},$$

from which (ii) follows, as before, by a combination of (**H4**) and (D.6). $\qquad\square$

Proof of Theorem 4.10. Write

$$\phi_{0:n}^\varepsilon h_n - \phi_{0:n} h_n = \sum_{k=0}^{n-1} \left(\phi_{0:n}^\varepsilon \bar{h}_{k|n} - \phi_{0:n} \bar{h}_{k|n} \right),$$

where each term can be decomposed according to

$$\phi_{0:n}^\varepsilon \bar{h}_{k|n} - \phi_{0:n} \bar{h}_{k|n} = \sum_{m=1}^{n} \left(\frac{\phi_{0:m}^\varepsilon \mathbf{L}_{m,n-1} \bar{h}_{k|n}}{\phi_{0:m}^\varepsilon \mathbf{L}_{m,n-1} \mathbb{1}_{\mathsf{X}^n}} - \frac{\phi_{0:m-1}^\varepsilon \mathbf{L}_{m-1,n-1} \bar{h}_{k|n}}{\phi_{0:m-1}^\varepsilon \mathbf{L}_{m-1,n-1} \mathbb{1}_{\mathsf{X}^n}} \right)$$

(recall that $\phi_0^\varepsilon \propto \chi$). In order to bound each term of this decomposition, write, using Lemma D.1,

$$\frac{\phi_{0:m}^\varepsilon \mathbf{L}_{m,n-1} \bar{h}_{k|n}}{\phi_{0:m}^\varepsilon \mathbf{L}_{m,n-1} \mathbb{1}_{\mathsf{X}^n}} - \frac{\phi_{0:m-1}^\varepsilon \mathbf{L}_{m-1,n-1} \bar{h}_{k|n}}{\phi_{0:m-1}^\varepsilon \mathbf{L}_{m-1,n-1} \mathbb{1}_{\mathsf{X}^n}} = \frac{\phi_m^\varepsilon \boldsymbol{\mathcal{L}}_{m,n}^\varepsilon \bar{h}_{k|n}}{\phi_m^\varepsilon \boldsymbol{\mathcal{L}}_{m,n}^\varepsilon \mathbb{1}_{\mathsf{X}^n}} - \frac{\phi_{m-1}^\varepsilon \boldsymbol{\mathcal{L}}_{m-1,n}^\varepsilon \bar{h}_{k|n}}{\phi_{m-1}^\varepsilon \boldsymbol{\mathcal{L}}_{m-1,n}^\varepsilon \mathbb{1}_{\mathsf{X}^n}}.$$

Now, for all $m \in [\![1, n]\!]$, pick an arbitrary element $x_m^* \in \mathsf{X}_m$ and define the kernel

$$\bar{\boldsymbol{\mathcal{L}}}_{m,n}^\varepsilon h(x_m) := \frac{\boldsymbol{\mathcal{L}}_{m,n}^\varepsilon h(x_m)}{\boldsymbol{\mathcal{L}}_{m,n}^\varepsilon \mathbb{1}_{\mathsf{X}^n}(x_m)} - \frac{\boldsymbol{\mathcal{L}}_{m,n}^\varepsilon h(x_m^*)}{\boldsymbol{\mathcal{L}}_{m,n}^\varepsilon \mathbb{1}_{\mathsf{X}^n}(x_m^*)}, \quad x_m \in \mathsf{X}_m, \quad h \in \mathsf{F}(\mathcal{X}^n). \tag{D.8}$$

Combining this definition with Lemma D.2, we may express the quantity of interest as

$$\phi_{0:n}^\varepsilon \bar{h}_{k|n} - \phi_{0:n} \bar{h}_{k|n} = \sum_{m=1}^{k} \left(\varphi_{m,n}^\varepsilon \bar{\boldsymbol{\mathcal{L}}}_{m,n}^\varepsilon \bar{h}_{k|n} - \breve{\varphi}_{m,n}^\varepsilon \bar{\boldsymbol{\mathcal{L}}}_{m,n}^\varepsilon \bar{h}_{k|n} \right)$$

$$+ \frac{\phi_{k+1}^\varepsilon \boldsymbol{\mathcal{L}}_{k+1,n}^\varepsilon \bar{h}_{k|n}}{\phi_{k+1}^\varepsilon \boldsymbol{\mathcal{L}}_{k+1,n}^\varepsilon \mathbb{1}_{\mathsf{X}^n}} - \frac{\phi_k^\varepsilon \boldsymbol{\mathcal{L}}_{k,n}^\varepsilon \bar{h}_{k|n}}{\phi_k^\varepsilon \boldsymbol{\mathcal{L}}_{k,n}^\varepsilon \mathbb{1}_{\mathsf{X}^n}} + \sum_{m=k+1}^{n-1} \left(\vec{\varphi}_{m,n}^\varepsilon \bar{\boldsymbol{\mathcal{L}}}_{m,n}^\varepsilon \bar{h}_{k|n} - \varphi_{m,n}^\varepsilon \bar{\boldsymbol{\mathcal{L}}}_{m,n}^\varepsilon \bar{h}_{k|n} \right).$$

Now, applying Lemmas D.3 and D.4 to the previous decomposition yields

$$\left| \phi_{0:n}^\varepsilon \bar{h}_{k|n} - \phi_{0:n} \bar{h}_{k|n} \right| \leq 2c\varepsilon \frac{\sigma_+}{\sigma_-^2} \sum_{k=0}^{n-1} \|\tilde{h}_k\|_\infty \left(\sum_{m=1}^{n-1} \rho^{|k-m|-1} + 1 \right).$$

which was to be established.

The second inequality follows straightforwardly according to

$$\sum_{k=0}^{n-1} \|\tilde{h}_k\|_\infty \left(\sum_{m=1}^{n-1} \rho^{|k-m|-1} + 1 \right) \leq \left(n + \frac{1}{\rho} \left(n + 2 \sum_{\ell=1}^{n-1} (n-\ell) \rho^\ell \right) \right) \sup_{k \in [\![0, n-1]\!]} \|\tilde{h}_k\|_\infty$$

$$\leq n \left(1 + \frac{1}{\rho} + \frac{2}{1-\rho} \right) \sup_{k \in [\![0, n-1]\!]} \|\tilde{h}_k\|_\infty.$$

$\square$

Appendix E: Proofs of Theorem 4.6, Corollary 4.7, and Proposition 4.9

Proof of Theorem 4.6. The following is a refinement of the proof of [29, Proposition 7].

We start bounding the first term of (4.6). For this purpose, we first prove that for all $m \in [\![0, n-1]\!]$,

$$\|\bar{\boldsymbol{\mathcal{R}}}^{\varepsilon}_{m,n} h_n\|_{\infty} \leq \|\mathbf{L}^{\varepsilon}_{m,n-1} \mathbb{1}_{\mathsf{X}^n}\|_{\infty} \sum_{k=0}^{n-1} \|\tilde{h}_k\|_{\infty} \rho^{|k-m|-1}. \tag{E.1}$$

In order to establish (E.1), first note that

$$\phi^{\varepsilon}_{0:n} h_n = \frac{\phi^{\varepsilon}_m \boldsymbol{\mathcal{R}}^{\varepsilon}_{m,n} h_n}{\phi^{\varepsilon}_m \boldsymbol{\mathcal{R}}^{\varepsilon}_{m,n} \mathbb{1}_{\mathsf{X}^n}}. \tag{E.2}$$

Thus, using the extensions (D.3) we may write, for $x_m \in \mathsf{X}_m$,

$$\frac{\bar{\boldsymbol{\mathcal{R}}}^{\varepsilon}_{m,n} h_n(x_m)}{\mathbf{L}^{\varepsilon}_{m,n-1} \mathbb{1}_{\mathsf{X}^n}(x_m)} = \sum_{k=0}^{n-1} \left(\frac{\boldsymbol{\mathcal{R}}^{\varepsilon}_{m,n} \bar{h}_{k|n}(x_m)}{\boldsymbol{\mathcal{R}}^{\varepsilon}_{m,n} \mathbb{1}_{\mathsf{X}^n}(x_m)} - \frac{\phi^{\varepsilon}_m \boldsymbol{\mathcal{R}}^{\varepsilon}_{m,n} \bar{h}_{k|n}}{\phi^{\varepsilon}_m \boldsymbol{\mathcal{R}}^{\varepsilon}_{m,n} \mathbb{1}_{\mathsf{X}^n}} \right).$$

Thus, the bound (E.1) is obtained by applying Lemma D.3 (twice: first, with $\varepsilon = 0$, allowing $\boldsymbol{\mathcal{L}}^{\varepsilon}_{m,n}$ in Lemma D.3 to be replaced by $\boldsymbol{\mathcal{L}}_{m,n}$; second, with the skew model playing the role of the original model, allowing $\boldsymbol{\mathcal{L}}_{m,n}$ to be replaced by $\boldsymbol{\mathcal{R}}^{\varepsilon}_{m,n}$), with $\lambda = \delta_{x_m}$ and $\lambda' = \phi^{\varepsilon}_m$, to each term on the right-hand side of the previous identity.

Now, using (E.1) and the bound (D.6),

$$\sum_{m=0}^{n-1} \frac{\phi^{\varepsilon}_m \vartheta_m \phi^{\varepsilon}_m \mathbf{L}^{\varepsilon}_m (\varphi_m [\bar{\boldsymbol{\mathcal{R}}}^{\varepsilon}_{m+1,n} h_n]^2)}{(\phi^{\varepsilon}_m \mathbf{L}^{\varepsilon}_{m,n-1} \mathbb{1}_{\mathsf{X}^n})^2}$$

$$\leq \sup_{\ell \in \mathbb{N}} \|\vartheta_\ell\|_{\infty} \sup_{\ell \in \mathbb{N}} \|\varphi_\ell\|_{\infty} \frac{\sigma^3_+}{\sigma^4_-} \sum_{m=0}^{n-1} \left(\sum_{k=0}^{n-1} \|\tilde{h}_k\|_{\infty} \rho^{|k-m-1|-1} \right)^2. \tag{E.3}$$

Then, since

$$\sum_{m=0}^{n-1} \left(\sum_{k=0}^{n-1} \rho^{|k-m-1|-1} \right)^2 \leq \frac{4n}{\rho^2 (1-\rho)^2},$$

we obtain the linear (in n) bound

$$\sum_{m=0}^{n-1} \frac{\phi^{\varepsilon}_m \vartheta_m \phi^{\varepsilon}_m \mathbf{L}^{\varepsilon}_m (\varphi_m [\bar{\boldsymbol{\mathcal{R}}}^{\varepsilon}_{m+1,n} h_n]^2)}{(\phi^{\varepsilon}_m \mathbf{L}^{\varepsilon}_{m,n-1} \mathbb{1}_{\mathsf{X}^n})^2}$$

$$\leq \sup_{\ell \in \mathbb{N}} \|\tilde{h}_\ell\|^2_{\infty} \sup_{\ell \in \mathbb{N}} \|\vartheta_\ell\|_{\infty} \sup_{\ell \in \mathbb{N}} \|\varphi_\ell\|_{\infty} \frac{4n}{\rho^2 (1-\rho)^5 \sigma_-}. \tag{E.4}$$

We turn to the second term of (4.6) and reusing (D.6) yields

$$\frac{\mathbf{L}_{\ell+1,m}^{\varepsilon}(\mathbf{B}_m^{\varepsilon}\varphi_m[\mathbf{L}_{m+1,n-1}^{\varepsilon}\mathbb{1}_{\mathsf{X}^n}]^2)}{(\phi_\ell^{\varepsilon}\mathbf{L}_{\ell,m-1}^{\varepsilon}\mathbb{1}_{\mathsf{X}^m})(\phi_m^{\varepsilon}\mathbf{L}_{m,n-1}^{\varepsilon}\mathbb{1}_{\mathsf{X}^n})^2} \leq \frac{\|\mathbf{L}_{\ell+1,m-1}^{\varepsilon}\mathbb{1}_{\mathsf{X}^m}\|_\infty\|\mathbf{L}_{m+1,n-1}^{\varepsilon}\mathbb{1}_{\mathsf{X}^n}\|_\infty^2\|\varphi_m\|_\infty\sigma_+}{(\phi_\ell^{\varepsilon}\mathbf{L}_{\ell,m-1}^{\varepsilon}\mathbb{1}_{\mathsf{X}^m})(\phi_m^{\varepsilon}\mathbf{L}_{m,n-1}^{\varepsilon}\mathbb{1}_{\mathsf{X}^n})^2}$$

$$\leq \frac{1}{(1-\rho)^4\sigma_-^2}\|\varphi_m\|_\infty. \tag{E.5}$$

By (2.13) it holds that

$$\mathbf{T}_\ell^{\varepsilon}h_\ell(x_\ell) + \tilde{h}_\ell(x_\ell, x_{\ell+1}) - \mathbf{T}_{\ell+1}^{\varepsilon}h_{\ell+1}(x_{\ell+1})$$
$$= \mathbf{T}_\ell^{\varepsilon}h_\ell(x_\ell) - \mathbf{B}_\ell^{\varepsilon}\mathbf{T}_\ell^{\varepsilon}h_\ell(x_{\ell+1}) + \tilde{h}_\ell(x_\ell, x_{\ell+1}) - \mathbf{B}_\ell^{\varepsilon}\tilde{h}_\ell(x_{\ell+1})$$
$$= \sum_{k=0}^{\ell-1}\left(\boldsymbol{\mathcal{R}}_{\ell,\ell}^{\varepsilon}\bar{h}_{k|\ell}(x_\ell) - \mathbf{B}_\ell^{\varepsilon}\boldsymbol{\mathcal{R}}_{\ell,\ell}^{\varepsilon}\bar{h}_{k|\ell}(x_{\ell+1})\right) + \tilde{h}_\ell(x_\ell, x_{\ell+1}) - \mathbf{B}_\ell^{\varepsilon}\tilde{h}_\ell(x_{\ell+1}),$$

and by applying Lemma D.3 (with $\lambda = \delta_{x_\ell}$ and $\lambda' = \mathbf{B}_\ell^{\varepsilon}(x_{\ell+1}, \cdot)$) to each term in the sum yields the bound

$$\|\mathbf{T}_\ell^{\varepsilon}h_\ell + \tilde{h}_\ell - \mathbf{T}_{\ell+1}^{\varepsilon}h_{\ell+1}\|_\infty \leq \left(\frac{1}{1-\rho} + 2\right)\sup_{\ell\in\mathbb{N}}\|\tilde{h}_\ell\|_\infty. \tag{E.6}$$

By combining (E.5) and (E.6) we obtain

$$\sum_{m=0}^{n-1}\sum_{\ell=0}^{m}\frac{\phi_m^{\varepsilon}\vartheta_m\phi_\ell^{\varepsilon}\mathbf{L}_\ell^{\varepsilon}\{\mathbf{B}_\ell^{\varepsilon}(\mathbf{T}_\ell^{\varepsilon}h_\ell + \tilde{h}_\ell - \mathbf{T}_{\ell+1}^{\varepsilon}h_{\ell+1})^2\mathbf{L}_{\ell+1,m}^{\varepsilon}(\mathbf{B}_m^{\varepsilon}\varphi_m[\mathbf{L}_{m+1,n-1}^{\varepsilon}\mathbb{1}_{\mathsf{X}^n}]^2)\}}{M^{m-\ell+1}(\phi_\ell^{\varepsilon}\mathbf{L}_{\ell,m-1}^{\varepsilon}\mathbb{1}_{\mathsf{X}^m})(\phi_m^{\varepsilon}\mathbf{L}_{m,n-1}^{\varepsilon}\mathbb{1}_{\mathsf{X}^n})^2}$$
$$\leq \frac{1}{(1-\rho)^4\sigma_-^2}\left(\frac{1}{1-\rho} + 2\right)^2\sup_{\ell\in\mathbb{N}}\|\tilde{h}_\ell\|_\infty^2\sup_{\ell\in\mathbb{N}}\|\vartheta_\ell\|_\infty\sup_{\ell\in\mathbb{N}}\|\varphi_\ell\|_\infty\sum_{m=0}^{n-1}\sum_{\ell=0}^{m}M^{\ell-(m+1)}. \tag{E.7}$$

Finally, since, for $M \geq 2$,

$$\lim_{m\to\infty}\sum_{\ell=0}^{m}M^{\ell-(m+1)} = \frac{1}{M-1},$$

taking the Cesáro mean provides

$$\limsup_{n\to\infty}\frac{1}{n}\sum_{m=0}^{n-1}\sum_{\ell=0}^{m}\frac{\phi_m^{\varepsilon}\vartheta_m\phi_\ell^{\varepsilon}\mathbf{L}_\ell^{\varepsilon}\{\mathbf{B}_\ell^{\varepsilon}(\mathbf{T}_\ell^{\varepsilon}h_\ell + \tilde{h}_\ell - \mathbf{T}_{\ell+1}^{\varepsilon}h_{\ell+1})^2\mathbf{L}_{\ell+1,m}^{\varepsilon}(\mathbf{B}_m^{\varepsilon}\varphi_m[\mathbf{L}_{m+1,n-1}^{\varepsilon}\mathbb{1}_{\mathsf{X}^n}]^2)\}}{M^{m-\ell+1}(\phi_\ell^{\varepsilon}\mathbf{L}_{\ell,m-1}^{\varepsilon}\mathbb{1}_{\mathsf{X}^m})(\phi_m^{\varepsilon}\mathbf{L}_{m,n-1}^{\varepsilon}\mathbb{1}_{\mathsf{X}^n})^2}$$
$$\leq \frac{1}{(M-1)(1-\rho)^4\sigma_-^2}\left(\frac{1}{1-\rho} + 2\right)^2\sup_{\ell\in\mathbb{N}}\|\tilde{h}_\ell\|_\infty^2\sup_{\ell\in\mathbb{N}}\|\vartheta_\ell\|_\infty\sup_{\ell\in\mathbb{N}}\|\varphi_\ell\|_\infty. \tag{E.8}$$

Finally, we complete the proof by combining (E.4) and (E.8). $\square$

Proof of Corollary 4.7. First, note that by (E.1),

$$\|\bar{\mathcal{R}}_{0,n}^{\varepsilon} h_n\|_{\infty} \leq \|\mathbf{L}_{0,n-1}^{\varepsilon} \mathbb{1}_{\mathsf{X}^n}\|_{\infty} \sup_{\ell \in \mathbb{N}} \|\tilde{h}_{\ell}\|_{\infty} \frac{1}{\rho(1-\rho)}.$$

Moreover, since for all $x_0 \in \mathsf{X}_0$,

$$\sigma_- \mu_1 \mathbf{L}_{1,n-1}^{\varepsilon} \mathbb{1}_{\mathsf{X}^n} \leq \mathbf{L}_{0,n-1}^{\varepsilon} \mathbb{1}_{\mathsf{X}^n}(x_0) \leq \sigma_+ \mu_1 \mathbf{L}_{1,n-1}^{\varepsilon} \mathbb{1}_{\mathsf{X}^n}, \qquad (\text{E.9})$$

it holds that

$$\frac{\|\bar{\mathcal{R}}_{0,n}^{\varepsilon} h_n\|_{\infty}^2}{(\chi \mathbf{L}_{0,n-1}^{\varepsilon} \mathbb{1}_{\mathsf{X}^n})^2} \leq \sup_{\ell \in \mathbb{N}} \|\tilde{h}_{\ell}\|_{\infty}^2 \frac{1}{\rho^2 (1-\rho)^4 (\chi \mathbb{1}_{\mathsf{X}_0})^2}. \qquad (\text{E.10})$$

The desired bound is now completed by applying Theorem 4.6 and (E.10) to the two first and the last terms of $\sigma_n^2(h_n)$, respectively. $\qquad \square$

Proof of Proposition 4.9. The result follows straightforwardly from the bound (E.3); indeed, as the mapping $\mathsf{X}^n \ni x_{0:n} \mapsto f(x_n)$ belongs to $\mathsf{A}(\mathcal{X}^n)$, (E.3) implies that

$$\sum_{m=0}^{n-1} \frac{\phi_m^{\varepsilon} \vartheta_m \phi_m^{\varepsilon} \mathbf{L}_m^{\varepsilon}(\varphi_m[\mathbf{L}_{m+1}^{\varepsilon} \cdots \mathbf{L}_{n-1}^{\varepsilon}(f - \phi_n^{\varepsilon} f)]^2)}{(\phi_m^{\varepsilon} \mathbf{L}_m^{\varepsilon} \cdots \mathbf{L}_{n-1}^{\varepsilon} \mathbb{1}_{\mathsf{X}_n})^2}$$

$$\leq \sup_{\ell \in \mathbb{N}} \|\vartheta_{\ell}\|_{\infty} \sup_{\ell \in \mathbb{N}} \|\varphi_{\ell}\|_{\infty} \|f\|_{\infty}^2 \frac{\sigma_+^3}{\sigma_-^4} \rho^{-2} \sum_{m=0}^{n-1} \rho^{2|n-m-2|},$$

where the sum on the right-hand side can be bounded by $1/(1 - \rho^2) + \rho^2$. In addition, by (E.1),

$$\|\mathbf{L}_0^{\varepsilon} \cdots \mathbf{L}_{n-1}^{\varepsilon}(f - \phi_n^{\varepsilon} f)\|_{\infty} \leq \|\mathbf{L}_0^{\varepsilon} \cdots \mathbf{L}_{n-1}^{\varepsilon} \mathbb{1}_{\mathsf{X}^n}\|_{\infty} \|f\|_{\infty} \rho^{n-2},$$

and reusing (E.9) yields

$$\frac{\|\mathbf{L}_0^{\varepsilon} \cdots \mathbf{L}_{n-1}^{\varepsilon}(f - \phi_n^{\varepsilon} f)\|_{\infty}^2}{(\chi \mathbf{L}_0^{\varepsilon} \cdots \mathbf{L}_{n-1}^{\varepsilon} \mathbb{1}_{\mathsf{X}_n})^2} \leq \|f\|_{\infty}^2 \frac{\rho^{2n}}{\rho^4 (1-\rho)^2 (\chi \mathbb{1}_{\mathsf{X}_0})^2}.$$

The proof is completed by combining these bounds. $\qquad \square$

Appendix F: A technical lemma

The following technical lemma is a straightforward adaption of [28, Lemma 14] to the framework of Section 2.

Lemma F.1. *Assume (**H1**–**2**) and let $\boldsymbol{\Psi}$ be some possibly unnormalised transition kernel on $\mathsf{X}_n \times \mathcal{X}_{n+1}$ having transition density in $\mathsf{F}(\mathcal{X}_n \otimes \mathcal{X}_{n+1})$ with respect to some reference measure on $\mathcal{X}_{n+1}$. Moreover, let $(\varphi_N)_{N \in \mathbb{N}}$ be a sequence of functions in $\mathsf{F}(\mathcal{X}_{n+1})$ for which*

(i) there exists $\varphi \in \mathsf{F}(\mathcal{X}_{n+1})$ such that for all $x \in \mathsf{X}_{n+1}$, $\lim_{N\to\infty} \varphi_N(x) = \varphi(x)$ $\mathbb{P}$-a.s. and

(ii) there exists $c \in \mathbb{R}_{>0}$ such that $\|\varphi\|_\infty \le c$ for all $N \in \mathbb{N}$.

Then $\phi_n^N \mathbf{\Psi} \varphi_N \xrightarrow{\mathbb{P}} \phi_n \mathbf{\Psi} \varphi$ as $N \to \infty$.