LIMIT THEOREMS FOR SEQUENTIAL MCMC METHODS

AXEL FINKE,* *National University of Singapore*
ARNAUD DOUCET,** *University of Oxford*
ADAM M. JOHANSEN,*** *University of Warwick & The Alan Turing Institute*

Abstract

Both sequential Monte Carlo (SMC) methods (a.k.a. 'particle filters') and sequential Markov chain Monte Carlo (sequential MCMC) methods constitute classes of algorithms which can be used to approximate expectations with respect to (a sequence of) probability distributions and their normalising constants. While SMC methods sample particles conditionally independently at each time step, sequential MCMC methods sample particles according to a Markov chain Monte Carlo (MCMC) kernel. Introduced over twenty years ago in [6], sequential MCMC methods have attracted renewed interest recently as they empirically outperform SMC methods in some applications. We establish an $\mathbb{L}_r$-inequality (which implies a strong law of large numbers) and a central limit theorem for sequential MCMC methods and provide conditions under which errors can be controlled uniformly in time. In the context of state-space models, we also provide conditions under which sequential MCMC methods can indeed outperform standard SMC methods in terms of asymptotic variance of the corresponding Monte Carlo estimators.

Keywords: Sequential Monte Carlo methods; particle filters; Markov chain Monte Carlo methods; almost sure convergence; multivariate central limit theorem; time-uniform convergence

2010 Mathematics Subject Classification: Primary 65C05; 60F05; 60F25; 60G35
Secondary 60J05; 62F15; 68U20

1. Introduction

Sequential Monte Carlo (SMC) algorithms are used to approximate expectations with respect to a sequence of probability measures as well as the normalising constants of those measures. These techniques have found numerous applications in statistics, signal processing, physics, and related fields (see, e.g., [21] for a recent review). These algorithms proceed in a sequential manner by generating a collection of N conditionally independent particles at each time step. An alternative to these schemes in which the particles at each time step are sampled instead according to a single Markov chain Monte Carlo (MCMC) chain was proposed early on by [6]. Over recent years, there has been a renewed interest in such ideas as there is empirical evidence that these methods can outperform standard SMC algorithms in interesting scenarios (see, e.g., [8], [15], [28] ,[31], and [32] for novel applications and extensions). These methods have been termed *sequential MCMC* methods in the literature. However, in this work, we will

also refer to these as *MCMC particle filters (MCMC-PFs)*, to convey the idea that they rely on the same importance-sampling construction as particle methods.

Although there is a wealth of theoretical results available for SMC algorithms—see, for example, [10]—to the best of our knowledge, no convergence guarantees have yet been provided for MCMC-PFs. The present work fills this gap by providing an $\mathbb{L}_r$-inequality (which implies a strong law of large numbers) and a central limit theorem for the Monte Carlo estimators of expectations and normalising constants obtained through MCMC-PFs. Our results show that compared to conventional particle filters (PFs), the asymptotic variance of estimators obtained by MCMC-PFs includes additional terms which can be identified as the excess variance arising from the autocorrelation of the MCMC chains used to generate the particles. This implies that a standard PF always provides estimators with a lower asymptotic variance than the corresponding MCMC-PF if both algorithms target the same distributions and if the latter relies on MCMC kernels which are positive, i.e. which induce positive autocorrelations of all orders.

However, MCMC-PFs exhibit a significant advantage over regular PFs. The popular fully-adapted auxiliary particle filter (FA-APF) introduced by [29] typically significantly outperforms the bootstrap particle filter (BPF) of [16], for example when approximating the optimal filter for state-space models in the presence of informative measurements. Unfortunately, the FA-APF is implementable for only a very restricted class of models, whereas the MCMC-PF version of the FA-APF is much more widely applicable. In scenarios in which the FA-APF is not implementable, but its MCMC-PF version is, and in which the MCMC kernels used by the latter are sufficiently rapidly mixing, the MCMC-PF can substantially outperform implementable but rather inefficient standard PFs such as the BPF.

2. MCMC-PFs

2.1. Notation

Let $(\Omega, \mathcal{A}, \mathbb{P})$ be some probability space and denote expectation with respect to $\mathbb{P}$ by $\mathbb{E}$. For some set measurable space $(H, \mathcal{H})$, we let $\mathcal{B}(H)$ denote the Banach space of all bounded, real-valued, $\mathcal{H}$-measurable functions on H, equipped with the uniform norm $\|f\| := \sup_{x \in H} |f(x)|$. We also endow this space with the Borel σ-algebra (with respect to $\|\cdot\|$), and the product spaces $\mathcal{B}(H) \times \mathcal{B}(H)$ and $\mathcal{B}(H)^d$ for $d \in \mathbb{N}$ with the associated product σ-algebras. We also define the subsets $\mathcal{B}_1(H) := \{f \in \mathcal{B}(H) \mid \|f\| \le 1\}$. Finally, we let $\mathbf{1} \in \mathcal{B}(H)$ denote the unit function on H, i.e. $\mathbf{1} \equiv 1$.

Let $\mathcal{M}(H)$ denote the Banach space of all finite and signed measures on $(H, \mathcal{H})$ equipped with the total variation norm $\|\mu\| := \frac{1}{2} \sup_{f \in \mathcal{B}_1(H)} |\mu(f)|$, where we use the shorthand $\mu(f) := \int_H f(x)\mu(\mathrm{d}x)$, for any $\mu \in \mathcal{M}(H)$ and any $f \in \mathcal{B}(H)$. We define $\mathcal{P}(H) \subset \mathcal{M}(H)$ to be the set of all probability measures on $(H, \mathcal{H})$. For any $\nu \in \mathcal{P}(H)$, and any $f, g \in \mathcal{B}(H)$, we further define $\mathrm{cov}_\nu[f, g] := \nu([f - \nu(f)][g - \nu(g)])$ and $\mathrm{var}_\nu[f] := \mathrm{cov}_\nu[f, f]$.

Let $(H', \mathcal{H}')$ be another measurable space. For any bounded integral operator $M : \mathcal{B}(H') \to \mathcal{B}(H)$, defined by $f \mapsto M(f)(x) := \int_{H'} f(z)M(x, \mathrm{d}z)$ for any $x \in H$, we define $[\mu \otimes M](f) = \int_{H \times H'} \mu(\mathrm{d}x)M(x, \mathrm{d}y)f(x, y)$ for any μ in $\mathcal{P}(H)$ and $f \in \mathcal{B}(H \times H')$. We also define the operator norm $\|M\| := \sup_{f \in \mathcal{B}_1(H')} \|M(f)\|$ as well as the *Dobrushin coefficient*: $\beta(M) := \sup_{(x,y) \in H' \times H'} \|M(x, \cdot) - M(y, \cdot)\|$.

Finally, '$\to_{\mathrm{a.s.}}$' denotes almost sure convergence with respect to $\mathbb{P}$ and '$\to_d$' denotes convergence in distribution.

2.2. Path-space Feynman–Kac model

We want to approximate expectations under some distributions which are related to a distribution flow $(\eta_n)_{n \geq 1}$ on spaces $(\mathbf{E}_n, \mathcal{E}_n)$—with $(\mathbf{E}_1, \mathcal{E}_1) := (E, \mathcal{E})$ and $(\mathbf{E}_n, \mathcal{E}_n) := (\mathbf{E}_{n-1} \times E, \mathcal{E}_{n-1} \otimes \mathcal{E})$, for $n > 1$—of increasing dimension, where

$$\eta_n(\mathrm{d}\mathbf{x}_n) := \frac{\gamma_n(\mathrm{d}\mathbf{x}_n)}{\mathcal{Z}_n} \in \mathcal{P}(\mathbf{E}_n)$$

for some positive finite measure γ_n on $(\mathbf{E}_n, \mathcal{E}_n)$ and typically unknown normalising constant $\mathcal{Z}_n := \gamma_n(\mathbf{1})$. Throughout this work, we write $\mathbf{x}_p := x_{1:p} = (\mathbf{x}_{p-1}, x_p)$ and $\mathbf{z}_p := z_{1:p} = (\mathbf{z}_{p-1}, z_p)$.

We assume that the target distributions are induced by a Feynman–Kac model on the path space [10]. That is, there exist an initial distribution $M_1 := \eta_1 \in \mathcal{P}(\mathbf{E}_1)$, a sequence of Markov transition kernels $M_n \colon \mathbf{E}_{n-1} \times \mathcal{E} \to [0, 1]$ for $n > 1$, and a sequence of bounded (without loss of generality we take the bound to be 1) measurable potential functions $G_n \colon \mathbf{E}_n \to (0, 1]$ for $n \geq 1$, such that for any $f_n \in \mathcal{B}(\mathbf{E}_n)$,

$$\gamma_n(f_n) = \eta_1 Q_{1,n}(f_n),$$

where we have defined the two-parameter semigroup

$$Q_{p,q}(f_q)(\mathbf{x}_p) := \begin{cases} [Q_{p+1} \cdots Q_q](f_q)(\mathbf{x}_p) & \text{if } p < q, \\ f_p(\mathbf{x}_p) & \text{if } p = q, \end{cases}$$

for any $1 \leq p \leq q \leq n$, with

$$Q_{p+1}(\mathbf{x}_p, \mathrm{d}\mathbf{z}_{p+1}) := G_p(\mathbf{z}_p)\delta_{\mathbf{x}_p}(\mathrm{d}\mathbf{z}_p)M_{p+1}(\mathbf{z}_p, \mathrm{d}z_{p+1}).$$

This implies that $\gamma_1 = \eta_1$ and, for $n > 1$,

$$\eta_n(f_n) = \Phi_n^{\eta_{n-1}}(f_n) := \frac{\eta_{n-1}Q_n(f_n)}{\eta_{n-1}Q_n(\mathbf{1})} = \frac{\gamma_n(f_n)}{\gamma_n(\mathbf{1})}, \tag{1}$$

where we have defined the following family of probability measures, indexed by $\mu \in \mathcal{P}(\mathbf{E}_{n-1})$:

$$\Phi_n^\mu(\mathrm{d}\mathbf{x}_n) := \begin{cases} M_1(\mathrm{d}\mathbf{x}_1) = \eta_1(\mathrm{d}\mathbf{x}_1), & \text{if } n = 1, \\ \dfrac{G_{n-1}(\mathbf{x}_{n-1})}{\mu(G_{n-1})}[\mu \otimes M_n](\mathrm{d}\mathbf{x}_n), & \text{if } n > 1. \end{cases}$$

We note that Φ_1^μ does not actually depend on μ, but we keep μ in the superscript in order to avoid the need to treat the case $n = 1$ separately.

For later use, we also define the family of normalised operators

$$\overline{Q}_{p,n}(f_n)(\mathbf{x}_p) := \frac{Q_{p,n}(f_n)(\mathbf{x}_p)}{\eta_p Q_{p,n}(\mathbf{1})}.$$

Note that this implies that $\eta_n(f_n) = \eta_p \overline{Q}_{p,n}(f_n)$ for any $1 \leq p \leq n$ and any $f_n \in \mathcal{B}(\mathbf{E}_n)$.

2.3. Generic MCMC-PF algorithm

In Algorithm 1, we summarise a generic MCMC-PF scheme for constructing sampling approximations η_n^N of η_n. It admits all the MCMC-PF discussed in this work as special cases. This algorithm is essentially a PF in which the particles are not sampled conditionally independently from Φ_n^μ at step n, for some $\mu \in \mathcal{P}(\mathbf{E}_{n-1})$, but are instead sampled according to a Markov chain with initial distribution $\kappa_n^\mu \in \mathcal{P}(\mathbf{E}_n)$ and Markov transition kernels $K_n^\mu \colon \mathbf{E}_n \times \mathcal{E}_n \to [0, 1]$ which are invariant with respect to Φ_n^μ, i.e. $\Phi_n^\mu K_n^\mu = \Phi_n^\mu$. Recall that Φ_1^μ does not actually depend on μ (and we also assume that the same is true for κ_1^μ and K_1^μ), so that we need not define η_0^N.

Algorithm 1. (Generic MCMC-PF.) *At time $n \geq 1$,*

(1) *sample $\boldsymbol{\xi}_n^1 \sim \kappa_n^{\eta_{n-1}^N}$ and $\boldsymbol{\xi}_n^i \sim K_n^{\eta_{n-1}^N}(\boldsymbol{\xi}_n^{i-1}, \, \cdot \,)$, for $2 \leq i \leq N$;*

(2) *set $\eta_n^N := \frac{1}{N} \sum_{i=1}^N \delta_{\boldsymbol{\xi}_n^i}$.*

For any time $n \geq 1$ and for any $f_n \in \mathcal{B}(\mathbf{E}_n)$, $\gamma_n^N(f_n) := \eta_n^N(f_n) \prod_{p=1}^{n-1} \eta_p^N(G_p)$ is an estimate of $\gamma_n(f_n)$. In particular, an estimate of the normalising constant $\mathcal{Z}_n$ is

$$\mathcal{Z}_n^N := \gamma_n^N(\mathbf{1}) = \prod_{p=1}^{n-1} \eta_p^N(G_p) = \prod_{p=1}^{n-1} \frac{1}{N} \sum_{i=1}^N G_p(\boldsymbol{\xi}_p^i). \tag{2}$$

We hereafter write $\Phi_n^N := \Phi_n^{\eta_{n-1}^N}$, $\kappa_n^N := \kappa_n^{\eta_{n-1}^N}$, and $K_n^N := K_n^{\eta_{n-1}^N}$ to simplify the notation. Note that standard PFs are a special case of Algorithm 1 corresponding to $K_n^N(\mathbf{x}_n, \, \cdot \,) \equiv \Phi_n^N(\cdot) = \kappa_n^N(\cdot)$. Unfortunately, implementing standard PFs can become prohibitively costly whenever there is no cheap way of generating N independent and identically distributed (IID) samples from Φ_n^N—which can be the case when Φ_n^N is chosen for reasons of statistical efficiency rather than computational convenience, as in the case of the FA-APF of [29]. In contrast, Algorithm 1 only requires the construction of MCMC kernels which leave this distribution invariant.

Practitioners typically initialise the Markov chains close to stationarity by selecting $\kappa_n^N = [\eta_{n-1}^N \otimes M_n'](K_n^N)^{N_{\text{burnin}}}$ for some approximation $M_n'(\mathbf{x}_{n-1}, \mathrm{d}x_n)$ of $M_n(\mathbf{x}_{n-1}, \mathrm{d}x_n)$ (see, e.g., [8], [15] ,[31], [32]). Here, $N_{\text{burnin}} \geq 1$ denotes a suitably large number of iterations whose samples are discarded as 'burn-in'. Corollary 1, below, will demonstrate that, under regularity conditions, such algorithms can provide strongly consistent estimates of quantities of interest in spite of this out-of-equilibrium initialisation.

In situations in which it is possible to initialise the Markov chains at stationarity, i.e. in which we can initialise $\boldsymbol{\xi}_n^1 \sim \kappa_n^N = \Phi_n^N$, [14] showed that the estimator $\mathcal{Z}_n^N$ given in (2) is unbiased as for standard PFs [10]. This unbiasedness property permits—in principle—the use of MCMC-PFs within pseudo-marginal algorithms [3], and thus makes it possible to perform Bayesian parameter inference for state-space models. As such an initialisation only requires *one* draw from Φ_n^N, the use of relatively expensive methods, such as rejection sampling, may be justifiable. This is in contrast to standard PFs, which require N such draws, the cost of which may be prohibitive. However, as we discuss below in Subsections 2.5 and 4.2, the main advantage of MCMC-PFs over standard PFs is that the former can sometimes be implemented in scenarios in which the potential functions G_n cannot be evaluated and then the estimate $\mathcal{Z}_n^N$ from (2) cannot be evaluated either.

Furthermore, the conditional SMC scheme proposed in [1] can also be extended to MCMC-PFs as demonstrated in [33]. As discussed in [14], the resulting 'conditional' MCMC-PF can be used within the particle Gibbs sampler from [3]. Importantly, it requires neither the construction of a suitable initial distribution κ_n^N nor the ability to evaluate the potential functions G_n.

The literature on MCMC algorithms provides numerous ways in which to construct the Markov kernels K_n^μ. For instance, we could use Metropolis–Hastings (MH) [6], the Metropolis-adjusted Langevin algorithm, Hamiltonian Monte Carlo and hybrid kernels [31], [32], or kernels based on invertible particle flow ideas [22] or on the bouncy particle sampler [28]. As an illustration, Example 1 describes a simple independent MH kernel with a proposal distribution tailored to our setting.

Example 1. (*Independent MH.*) For any $n \geq 1$ and any $\mu \in \mathcal{P}(\mathbf{E}_{n-1})$, define the proposal distribution

$$\Pi_n^\mu(\mathrm{d}\mathbf{x}_n) := \begin{cases} R_1(\mathrm{d}\mathbf{x}_1) & \text{if } n = 1, \\ \dfrac{F_{n-1}(\mathbf{x}_{n-1})}{\mu(F_{n-1})}[\mu \otimes R_n](\mathrm{d}\mathbf{x}_n) & \text{if } n > 1, \end{cases}$$

for some sequence of nonnegative bounded measurable functions $F_n \colon \mathbf{E}_n \to [0, \infty)$, some distribution $R_1 \in \mathcal{P}(\mathbf{E}_1)$ with $M_1 \ll R_1$, and some sequence of Markov transition kernels $R_n \colon \mathbf{E}_{n-1} \times \mathcal{E} \to [0, 1]$ with $M_n(\mathbf{x}_{n-1}, \cdot) \ll R_n(\mathbf{x}_{n-1}, \cdot)$, for any $\mathbf{x}_{n-1} \in \mathbf{E}_{n-1}$; for any $n \geq 1$, both F_{n-1} and R_n are assumed to satisfy

$$\sup_{\mathbf{x}_n \in \mathbf{E}_n} \frac{G_{n-1}(\mathbf{x}_{n-1})}{F_{n-1}(\mathbf{x}_{n-1})} \frac{\mathrm{d}M_n(\mathbf{x}_{n-1}, \cdot)}{\mathrm{d}R_n(\mathbf{x}_{n-1}, \cdot)}(x_n) < \infty. \tag{3}$$

The independent MH kernel K_n^μ with proposal distribution Π_n^μ and target/invariant distribution Φ_n^μ is given by

$$K_n^\mu(\mathbf{x}_n, \mathrm{d}\mathbf{z}_n) := \alpha_n(\mathbf{x}_n, \mathbf{z}_n)\Pi_n^\mu(\mathrm{d}\mathbf{z}_n)$$
$$+ \left(1 - \int_{\mathbf{E}_n} \alpha_n(\mathbf{x}_n, \mathrm{d}\mathbf{x}_n')\Pi_n^\mu(\mathrm{d}\mathbf{x}_n')\right)\delta_{\mathbf{x}_n}(\mathrm{d}\mathbf{z}_n),$$

with acceptance probability

$$\alpha_n(\mathbf{x}_n, \mathbf{z}_n) := 1 \wedge \frac{\mathrm{d}\Phi_n^\mu}{\mathrm{d}\Pi_n^\mu}(\mathbf{z}_n) \Big/ \frac{\mathrm{d}\Phi_n^\mu}{\mathrm{d}\Pi_n^\mu}(\mathbf{x}_n)$$
$$= 1 \wedge \frac{\dfrac{G_{n-1}}{F_{n-1}}(\mathbf{z}_{n-1}) \dfrac{\mathrm{d}M_n(\mathbf{z}_{n-1}, \cdot)}{\mathrm{d}R_n(\mathbf{z}_{n-1}, \cdot)}(z_n)}{\dfrac{G_{n-1}}{F_{n-1}}(\mathbf{x}_{n-1}) \dfrac{\mathrm{d}M_n(\mathbf{x}_{n-1}, \cdot)}{\mathrm{d}R_n(\mathbf{x}_{n-1}, \cdot)}(x_n)}. \tag{4}$$

This acceptance probability notably does not depend on μ.

2.4. Computational cost

If we are interested only in approximating the normalising constant $\mathcal{Z}_n$, and if $G_{n-1}(\mathbf{x}_{n-1})$ and $M_n(\mathbf{x}_{n-1}, \cdot)$ depend upon only a fixed number of the most recent component(s) of $\mathbf{x}_{n-1}$

(as is the case in the state-space models discussed below), Algorithm 1 can be implemented at a per-time-step complexity (in both space and time) that is linear in the number of particles N and constant in the time horizon n.

2.5. Application to state-space models

Let $(F, \mathcal{F})$ be another measurable space. The MCMC-PF may be used for (but is not limited to) performing inference in a state-space model given by the bivariate Markov chain $(X_n, Y_n)_{n \geq 1}$ on $(E \times F, \mathcal{E} \vee \mathcal{F})$ with initial distribution $L_1(\mathrm{d}x_1)g_1(x_1, y_1)\psi(\mathrm{d}y_1)$ and with Markov transition kernels (for any $n > 1$)

$$L_n(x_{n-1}, \mathrm{d}x_n)g_n(x_n, y_n)\psi(\mathrm{d}y_n).$$

Here, $L_1 \in \mathcal{P}(E)$ is some initial distribution for X_1, and $L_n : E \times \mathcal{E} \to [0, 1]$, for $n > 1$, is a Markov transition kernel. Furthermore, ψ is some suitable σ-finite dominating measure on $(F, \mathcal{F})$ and some positive bounded function $g_n(\cdot, y_n)$ so that $g_n(x_n, y_n)\psi(\mathrm{d}y_n)$ represents the transition kernels for the observation at time n. Usually, we can only observe realisations of $(Y_n)_{n \geq 1}$, whereas the process $(X_n)_{n \geq 1}$ is latent.

Assume that we have observed realisations $\mathbf{y}_n = (y_1, \ldots, y_n)$ of $\mathbf{Y}_n := (Y_1, \ldots, Y_n)$; then we often wish to compute (expectations under) the

- *filter:* $\pi_n(f_n) := \mathbb{E}[f_n(\mathbf{X}_n)|\mathbf{Y}_n = \mathbf{y}_n]$, for $f_n \in \mathcal{B}(\mathbf{E}_n)$;
- *predictor:* $\tilde{\pi}_n(f_n) := \mathbb{E}[f_n(\mathbf{X}_n)|\mathbf{Y}_{n-1} = \mathbf{y}_{n-1}]$, for $f_n \in \mathcal{B}(\mathbf{E}_n)$;
- *marginal likelihood:* $\mathcal{L}_n := \mathbb{E}[\prod_{p=1}^{n} g_p(X_p, y_p)]$.

Note that the definitions of 'filter' and 'predictor' here refer to the historical process, as we are taking a path-space approach. These terms are sometimes reserved for the final-component marginals of π_n and $\tilde{\pi}_n$; we will use the terms *marginal filter* and *marginal predictor* for those objects.

Example 2. (*BPF-type flow.*) If, for any $n \geq 1$,

$$G_n(\mathbf{x}_n) := g_n(x_n, y_n),$$
$$M_n(\mathbf{x}_{n-1}, \mathrm{d}x_n) := L_n(x_{n-1}, \mathrm{d}x_n), \tag{5}$$

then $\eta_n = \tilde{\pi}_n$ is the time-n predictor, $\eta_n(G_n f_n)/\eta_n(G_n) = \pi_n(f_n)$ recovers the time-n filter, and $\mathcal{Z}_{n+1} = \mathcal{L}_n$ is the marginal likelihood associated with the observations $\mathbf{y}_n$ (with $\mathcal{Z}_1 = 1$).

In this case, Algorithm 1 can be implemented (e.g. using the independent MH kernel from Example 1) as long as g_n, F_n, and $\mathrm{d}L_n(x_{n-1}, \cdot)/\mathrm{d}R_n(\mathbf{x}_{n-1}, \cdot)$ can be evaluated pointwise.

Example 3. (*FA-APF-type flow.*) If, for any $n \geq 1$,

$$G_n(\mathbf{x}_n) := L_{n+1}(g_{n+1}(\cdot, y_{n+1}))(x_n), \tag{6}$$

$$M_n(\mathbf{x}_{n-1}, \mathrm{d}x_n) := \frac{L_n(x_{n-1}, \mathrm{d}x_n)g_n(x_n, y_n)}{L_n(g_n(\cdot, y_n))(x_{n-1})}, \tag{7}$$

then $\eta_n = \pi_n$ is the time-n filter, $\eta_{n-1} \otimes L_n = \tilde{\pi}_n$ recovers the time-n predictor, and $\mathcal{Z}_n = \mathcal{L}_n$ is the marginal likelihood associated with the observations $\mathbf{y}_n$.

For this flow, it follows from (1) that sampling $\boldsymbol{\xi}_n^i$ from Φ_n^N requires first sampling an index $J = j \in \{1, ..., N\}$ with probability proportional to $G_{n-1}(\boldsymbol{\xi}_{n-1}^j)$, setting the first $n-1$ components of $\boldsymbol{\xi}_n^i$ equal to $\boldsymbol{\xi}_{n-1}^j$, and then sampling the final component according to $M_n(\boldsymbol{\xi}_{n-1}^j, \cdot)$. There are many scenarios in which this is not feasible as both (6) and (7) involve an intractable integral. However, designing an MCMC kernel of invariant distribution Φ_n^N is a much easier task, as the product $G_{n-1}(\mathbf{x}_{n-1})M_n(\mathbf{x}_{n-1}, \mathrm{d}x_n)$ does not involve any intractable integral. For example, if we use the independent MH kernel from Example 1 then the acceptance probability in (4) reduces to the following (for simplicity, we take $F_{n-1} \equiv 1$):

$$\alpha_n(\mathbf{x}_n, \mathbf{z}_n) = 1 \wedge \frac{g_n(z_n, y_n)}{g_n(x_n, y_n)} \frac{\dfrac{\mathrm{d}L_n(z_{n-1}, \cdot)}{\mathrm{d}R_n(\mathbf{z}_{n-1}, \cdot)}(z_n)}{\dfrac{\mathrm{d}L_n(x_{n-1}, \cdot)}{\mathrm{d}R_n(\mathbf{x}_{n-1}, \cdot)}(x_n)}.$$

Example 4. (*General auxiliary particle filter (APF)-type flow.*) Let η_1 be some approximation of π_1. For $n \geq 1$, let $M_{n+1}(\mathbf{x}_n, \mathrm{d}x_{n+1})$ be some approximation of (7), and let

$$G_n(\mathbf{x}_n) := \begin{cases} \dfrac{\mathrm{d}L_1}{\mathrm{d}\eta_1}(x_1)g_1(x_1, y_1)\tilde{g}_1(x_1, y_2) & \text{if } n = 1, \\[2ex] \dfrac{\mathrm{d}L_n(x_{n-1}, \cdot)}{\mathrm{d}M_n(\mathbf{x}_{n-1}, \cdot)}(x_n)\dfrac{g_n(x_n, y_n)\tilde{g}_n(x_n, y_{n+1})}{\tilde{g}_{n-1}(x_{n-1}, y_n)} & \text{if } n > 1. \end{cases}$$

Here, $\tilde{g}_n(x_n, y_{n+1})$ denotes some tractable approximation of (6) which can be evaluated pointwise. More generally, we could incorporate information from observations at times $n+1, \ldots, n+l$ for some $l \geq 1$ into $M_{n+1}(\mathbf{x}_n, \mathrm{d}x_{n+1})$ and replace $\tilde{g}_n(x_n, y_{n+1})$ by some approximation of $\int_{E^l} \prod_{p=n+1}^{n+l} L_p(\mathbf{x}_{p-1}, \mathrm{d}x_p)g_p(x_p, y_p)$, as in the case of *lookahead* methods (see [23], for example).

Note that the (general) APF flow admits the two other flows as special cases. That is, taking M_n as in (5) and $\tilde{g}_n \equiv 1$ yields the BPF-type flow; taking M_n as in (7) and $\tilde{g}_n(x_n, y_{n+1}) = L_{n+1}(g_{n+1}(\cdot, y_{n+1}))(x_n)$ yields the FA-APF-type flow.

In the remainder of this work, we will refer to Algorithm 1 as the *MCMC bootstrap particle filter (MCMC-BPF)* whenever the distribution flow $(\eta_n)_{n\geq 1}$ is defined as in Example 2, as the *MCMC fully-adapted auxiliary particle filter (MCMC-FA-APF)* whenever the flow is defined as in Example 3, and as *MCMC auxiliary particle filters (MCMC-APF)* whenever the flow is defined as in Example 4. Furthermore, we drop the prefix 'MCMC' when referring to the conventional PF-analogues of these algorithms, i.e. in the case that $K_n^\mu(\mathbf{x}_n, \cdot) \equiv \Phi_n^\mu = \kappa_n^\mu$.

3. Main results

In this section, we state an $\mathbb{L}_r$-inequality (Proposition 1)—which also implies a strong law of large numbers (SLLN) (Corollary 1)—and a central limit theorem (CLT) (Proposition 2) for the approximations of the normalised and unnormalised flows $(\eta_n)_{n\geq 1}$ and $(\gamma_n)_{n\geq 1}$ generated by an MCMC-PF.

3.1. Assumptions

We make the following assumptions about the MCMC kernels used to sample the particles at each time step.

(A1) For any $n \geq 1$, there exists $i_n \in \mathbb{N}$ and $\varepsilon_n(K) \in (0, 1]$ such that for all $\mu \in \mathcal{P}(\mathbf{E}_{n-1})$ and all $\mathbf{x}_n, \mathbf{z}_n \in \mathbf{E}_n$,

$$(K_n^\mu)^{i_n}(\mathbf{x}_n, \; \cdot \;) \geq \varepsilon_n(K)(K_n^\mu)^{i_n}(\mathbf{z}_n, \; \cdot \;).$$

(A2) For any $n \geq 1$, there exists a constant $\overline{\Gamma}_n < \infty$ and a family of bounded integral operators $(\Gamma_n^\mu)_{\mu \in \mathcal{P}(\mathbf{E}_{n-1})}$ from $\mathcal{B}(\mathbf{E}_n)$ to $\mathcal{B}(\mathbf{E}_{n-1})$ such that for any $(\nu, \mu) \in \mathcal{P}(\mathbf{E}_{n-1})^2$ and any $f_n \in \mathcal{B}(\mathbf{E}_n)$,

$$\|[K_n^\nu - K_n^\mu](f_n)\| \leq \int_{\mathcal{B}(\mathbf{E}_{n-1})} |[\nu - \mu](g)| \Gamma_n^\mu(f_n, \mathrm{d}g)$$

and

$$\int_{\mathcal{B}(\mathbf{E}_{n-1})} \|g\| \Gamma_n^\mu(f_n, \mathrm{d}g) \leq \|f_n\| \overline{\Gamma}_n.$$

The first assumption on the MCMC kernels ensures that they are suitably ergodic (it corresponds to assuming that the kernels used are uniformly ergodic, uniformly in their invariant distribution) and is the only assumption required to obtain the $\mathbb{L}_r$-inequality and the SLLN. More precisely, recall that for any bounded integral operator $M \colon \mathcal{B}(H) \to \mathcal{B}(H)$, the associated Dobrushin coefficient is given by $\beta(M) := \sup_{(x,y) \in H \times H} \|M(x, \; \cdot \;) - M(y, \; \cdot \;)\|$. Note that Assumption **A1** implies that

$$\sup_{\mu \in \mathcal{P}(\mathbf{E}_{n-1})} \beta((K_n^\mu)^{i_n}) \leq 1 - \varepsilon_n(K) < 1.$$

In particular, if $(\boldsymbol{\xi}_p^i)_{i \geq 1}$ and $(\tilde{\boldsymbol{\xi}}_p^i)_{i \geq 1}$ are Markov chains with transition kernels K_n^μ, with $(\tilde{\boldsymbol{\xi}}_p^i)_{i \geq 1}$ initialised from stationarity, then a standard coupling argument shows that **A1** also implies that for any $N, r \geq 1$ and any $f_n \in \mathcal{B}(\mathbf{E}_n)$ with $\|f_n\| \leq 1$,

$$\mathbb{E}\left[\left|\sum_{i=1}^N f_n(\boldsymbol{\xi}_n^i) - f_n(\tilde{\boldsymbol{\xi}}_n^i)\right|^r\right]^{\frac{1}{r}} \leq \sum_{i=1}^N \mathbb{E}\left[\left|f_n(\boldsymbol{\xi}_n^i) - f_n(\tilde{\boldsymbol{\xi}}_n^i)\right|^r\right]^{\frac{1}{r}}$$

$$\leq 2\|f_n\| \sum_{i=1}^N (1 - \varepsilon_n(K))^{\lfloor i/i_n \rfloor}$$

$$\leq 2 i_n / \varepsilon_n(K) =: \overline{T}_n. \tag{8}$$

The second assumption on the MCMC kernels is a local Lipschitz condition. As shown in Lemma 3, this assumption ensures that for any $r \geq 1$, $\sqrt{N}$-convergence of $[\eta_{n-1}^N - \eta_{n-1}](f_{n-1})$ to zero in $\mathbb{L}_r$ (for all $f_{n-1} \in \mathcal{B}(\mathbf{E}_{n-1})$) implies $\sqrt{N}$-convergence of $\|[K_n^{\eta_{n-1}^N} - K_n^{\eta_{n-1}}](f_n)\|$ to zero in $\mathbb{L}_r$. By extension, in the proof of the CLT, specifically in Lemma 4, this property enables us to conclude that the (suitably scaled) variance of the particle approximation converges to a well-defined limiting asymptotic variance as $N \to \infty$.

Assumptions **A1** and **A2** are similar to those imposed in [5]. They are strong and rarely hold for non-compact spaces. It might be possible to adopt weaker conditions such as those in [2], but this would involve substantially more technical and complicated proofs. As an illustration, we show that Assumptions **A1** and **A2** hold if we employ the independent MH kernels from Example 1, at least if E is finite.

Example 5. (*Independent MH, continued.*) Assumption **A1** is satisfied thanks to [26, Theorem 2.1]. To see this, note that by (3), for any $n \geq 1$ and any $\mu \in \mathcal{P}(\mathbf{E}_{n-1})$, since F_{n-1} is bounded and $G_{n-1} > 0$,

$$\sup_{\mathbf{x}_n \in \mathbf{E}_n} \frac{\mathrm{d}\Phi_n^\mu}{\mathrm{d}\Pi_n^\mu}(\mathbf{x}_n) \leq \frac{\|F_{n-1}\|}{\mu(G_{n-1})} \sup_{\mathbf{x}_n \in \mathbf{E}_n} \frac{G_{n-1}(\mathbf{x}_{n-1})}{F_{n-1}(\mathbf{x}_{n-1})} \frac{\mathrm{d}M_n(\mathbf{x}_{n-1}, \cdot)}{\mathrm{d}R_n(\mathbf{x}_{n-1}, \cdot)}(x_n) < \infty.$$

Assumption **A2** was proved for finite spaces E (and in the case $F_n \equiv 1$, $R_n = M_n$) in [5, Section 2].

When proving time-uniform convergence results, we also make the following assumptions on the mutation kernels and potential functions of the Feynman–Kac model. The first of these ensures that Assumption **A1** holds uniformly in time. The second and third of these constitute strong mixing conditions that have been extensively used in the analysis of SMC algorithms; although they can often be relaxed in similar settings, this comes at the cost of greatly complicating the analysis (see, e.g., [36], [12]).

(B1) $\bar{\iota} := \sup_{n \geq 1} i_n < \infty$ and $\varepsilon(K) := \inf_{n \geq 1} \varepsilon_n(K) > 0$.

(B2) There exist $m \in \mathbb{N}$ and $\varepsilon(M) \in (0, 1]$ such that for any $n \geq 1$, any $\mathbf{x}_n, \mathbf{z}_n \in \mathbf{E}_n$ and any $\varphi \in \mathcal{B}(E)$,

$$\int_{E^m} \left[\prod_{p=1}^m M_{n+p}(\mathbf{x}_{n+p-1}, \mathrm{d}x_{n+p}) \right] \varphi(x_{n+m})$$

$$\geq \varepsilon(M) \int_{E^m} \left[\prod_{p=1}^m M_{n+p}(\mathbf{z}_{n+p-1}, \mathrm{d}z_{n+p}) \right] \varphi(z_{n+m}).$$

(B3) There exist $l \in \mathbb{N}$ and $\varepsilon(G) \in (0, 1]$ such that for any $n \geq 1$ and any $\mathbf{x}_n, \mathbf{z}_n \in \mathbf{E}_n$,

$$G_n(\mathbf{x}_n) = G_n((\mathbf{z}_{n-l-1}, x_{((n-l) \vee 1):n})) \quad \text{and} \quad G_n(\mathbf{x}_n) \geq \varepsilon(G) G_n(\mathbf{z}_n).$$

Under these conditions, time-uniform bounds will be obtained when the test function under study has supremum norm of at most 1 and depends upon only its final coordinate marginal; i.e. we will restrict our attention to test functions $f_n \in \mathcal{B}_1^\star(\mathbf{E}_n)^d$, where $\mathcal{B}_1^\star(\mathbf{E}_n) := \{f_n' \in \mathcal{B}^\star(\mathbf{E}_n) \mid \|f_n'\| \leq 1\}$ with

$$\mathcal{B}^\star(\mathbf{E}_n) := \{\varphi \circ \zeta_n \mid \varphi \in \mathcal{B}(E)\}.$$

Here, for any $n \geq 1$, $\zeta_n : \mathbf{E}_n \to E$ denotes the canonical final-coordinate projection operator defined through $\mathbf{x}_n \mapsto \zeta_n(\mathbf{x}_n) := x_n$. In the state-space model context this corresponds, essentially, to considering the approximation of the marginal filter and predictor rather than their path-space analogues.

3.2. Strong law of large numbers

The first main result in this work is the $\mathbb{L}_r$-inequality given in Proposition 1, the proof of which will be given in Appendix C. As an immediate consequence, we obtain the SLLN stated in Corollary 1.

Proposition 1. ($\mathbb{L}_r$-inequality.) *Under Assumption **A1**, for each r, $n \geq 1$ there exist a_n, $b_r < \infty$ such that for any $f_n \in \mathcal{B}(\mathbf{E}_n)$ and any $N \geq 1$,*

$$\mathbb{E}\big[\big|[\eta_n^N - \eta_n](f_n)\big|^r\big]^{\frac{1}{r}} \leq \frac{a_n b_r}{\sqrt{N}} \|f_n\|. \tag{9}$$

*Under the additional Assumptions **B1**–**B3** and if $f_n \in \mathcal{B}_1^\star(\mathbf{E}_n)$, the right-hand side of (9) is bounded uniformly in time; i.e. there exists $a < \infty$ such that $\sup_{n \geq 1} a_n \leq a$.*

Corollary 1. (Strong law of large numbers.) *Under Assumption **A1**, for any n, $d \geq 1$ and $f_n \in \mathcal{B}(\mathbf{E}_n)^d$, as $N \to \infty$,*

(1) $\eta_n^N(f_n) \to_{\text{a.s.}} \eta_n(f_n)$;

(2) $\gamma_n^N(f_n) \to_{\text{a.s.}} \gamma_n(f_n)$.

Proof. Without loss of generality, we prove the result for scalar-valued test functions $f_n \in \mathcal{B}(\mathbf{E}_n)$. Part 1 is a direct consequence of Proposition 1, for some $r > 2$, using the Borel–Cantelli lemma together with Markov's inequality. Part 2 follows from Part 1 and boundedness of the potential functions G_p, i.e.

$$\gamma_n^N(f_n) = \eta_n^N(f_n) \prod_{p=1}^{n-1} \eta_p^N(G_p) \to_{\text{a.s.}} \eta_n(f_n) \prod_{p=1}^{n-1} \eta_p(G_p) = \gamma_n(f_n),$$

as $N \to \infty$. This completes the proof. $\qquad\blacksquare$

3.3. Central limit theorem

The second main result is Proposition 2, which adapts the usual CLT for SMC algorithms from [10, Propositions 9.4.1 & 9.4.2] to our setting. Its proof is given in Appendix C. As in [11], [5], we will make extensive use of the resolvent operators T_n^μ, defined, for any $\mu \in \mathcal{P}(\mathbf{E}_{n-1})$ and any $f_n \in \mathcal{B}(\mathbf{E}_n)$, by

$$T_n^\mu(f_n) := \sum_{j=0}^{\infty} [(K_n^\mu)^j - \Phi_n^\mu](f_n).$$

These operators satisfy the *Poisson equation*

$$(K_n^\mu - \text{Id})T_n^\mu = \Phi_n^\mu - \text{Id}, \tag{10}$$

$$\Phi_n^\mu T_n^\mu \equiv 0. \tag{11}$$

Under Assumption **A1**, [5, Proposition 3.1] shows that, for $\overline{T}_n$ as defined in (8),

$$\sup_{\mu \in \mathcal{P}(\mathbf{E}_{n-1})} \|T_n^\mu\| \leq \overline{T}_n. \tag{12}$$

In the following, for $n \geq 1$, we consider a vector-valued test function $f_n = (f_n^u)_{1 \leq u \leq d} \in \mathcal{B}(\mathbf{E}_n)^d$. Using the resolvent operators, for any $1 \leq u$, $v \leq d$, we define the covariance function $C_n^\mu(f_n^u, f_n^v) \colon \mathbf{E}_n \to \mathbb{R}$ by

$$C_n^\mu(f_n^u, f_n^v)(\mathbf{x}_n)$$
$$:= K_n^\mu[(T_n^\mu(f_n^u) - K_n^\mu T_n^\mu(f_n^u)(\mathbf{x}_n))(T_n^\mu(f_n^v) - K_n^\mu T_n^\mu(f_n^v)(\mathbf{x}_n))](\mathbf{x}_n) \tag{13}$$
$$= \text{cov}_{K_n^\mu(\mathbf{x}_n, \,\cdot\,)}[T_n^\mu(f_n^u), T_n^\mu(f_n^v)]$$

for any $\mathbf{x}_n \in \mathbf{E}_n$. Under Assumption **A1**, we have $C_n^\mu(f_n^u, f_n^v) \in \mathcal{B}(\mathbf{E}_n)$ for any $1 \le u, v \le d$ and any $\mu \in \mathcal{P}(\mathbf{E}_{n-1})$. Indeed, using (12), it is straightforward to check that

$$\sup_{\mu \in \mathcal{P}(\mathbf{E}_{n-1})} \|C_n^\mu(f_n^u, f_n^v)\| \le 4\overline{T}_n^2 \|f_n^u\| \|f_n^v\|. \tag{14}$$

Throughout the remainder of this work, let $V = (V_n)_{n \ge 1}$ be a sequence of independent and centred Gaussian fields with

$$\mathbb{E}[V_n(f_n^u)V_n(f_n^v)] = \eta_n C_n^{\eta_{n-1}}(f_n^u, f_n^v), \tag{15}$$

and define the (d,d)-matrix $\Sigma_n(f_n) := (\Sigma_n(f_n^u, f_n^v))_{1 \le u,v \le d}$ by

$$\Sigma_n(f_n^u, f_n^v) := \sum_{p=1}^n \mathbb{E}[V_n(\overline{Q}_{p,n}(f_n^u))V_n(\overline{Q}_{p,n}(f_n^v))], \tag{16}$$

for any $n \ge 1$, any $f_n = (f_n^u)_{1 \le u \le d} \in \mathcal{B}(\mathbf{E}_n)^d$, and any $1 \le u, v \le d$. Additionally, let $N(0, \Sigma)$ denote a (multivariate) centred Gaussian distribution with some covariance matrix Σ.

Proposition 2. (Central limit theorem.) *Under Assumptions **A1** and **A2**, for any $n, d \ge 1$ and any $f_n \in \mathcal{B}(\mathbf{E}_n)^d$, as $N \to \infty$,*

$$\frac{\sqrt{N}}{\gamma_n(\mathbf{1})}[\gamma_n^N - \gamma_n](f_n) \to_{\mathrm{d}} \sum_{p=1}^n V_p(\overline{Q}_{p,n}(f_n)) \sim N(0, \Sigma_n(f_n)),$$

and likewise, writing $\bar{f}_n := f_n - \eta_n(f_n)$,

$$\sqrt{N}[\eta_n^N - \eta_n](f_n) \to_{\mathrm{d}} \sum_{p=1}^n V_p(\overline{Q}_{p,n}(\bar{f}_n)) \sim N(0, \Sigma_n(\bar{f}_n)). \tag{17}$$

*Under the additional Assumptions **B1**–**B3** and if $f_n \in \mathcal{B}_1^\star(\mathbf{E}_n)^d$, the asymptotic variance in (17) is bounded uniformly in time, i.e. there exists $c < \infty$ such that*

$$\sup \Sigma_n(\bar{f}_n^u, \bar{f}_n^v) \le c, \tag{18}$$

where the supremum is over all $n \ge 1$, $f_n \in \mathcal{B}_1^\star(\mathbf{E}_n)^d$, and $1 \le u, v \le d$.

4. Comparison with standard PFs

4.1. Variance decomposition

In this section, we first examine the asymptotic variance from Proposition 2. We then illustrate the trade-off between MCMC-PFs and standard PFs.

To ease the exposition, we consider only scalar-valued test functions $f_n \in \mathcal{B}(\mathbf{E}_n)$ throughout this section. As noted in [5, Proposition 3.6], the terms $\Phi_n^\mu C_n^\mu(f_n, f_n)$ from (15), which, via (16), appear in the expressions for the asymptotic variance in Proposition 2, can be written in the following form, which is more commonly used in the MCMC literature:

$$\Phi_n^\mu C_n^\mu(f_n, f_n) = \int_{\mathbf{E}_n^2} \Phi_n^\mu(\mathrm{d}\mathbf{x}_n) K_n^\mu(\mathbf{x}_n, \mathrm{d}\mathbf{z}_n)\big[T_n^\mu(f_n)(\mathbf{z}_n) - K_n^\mu T_n^\mu(f_n)(\mathbf{x}_n)\big]^2$$

$$= \Phi_n^\mu\big(T_n^\mu(f_n)^2 - K_n^\mu T_n^\mu(f_n)^2\big)$$

$$= \Phi_n^\mu\big(T_n^\mu(f_n)^2 - [\Phi_n^\mu(f_n) - f_n + T_n^\mu(f_n)]^2\big) \quad \text{[by (10)]}$$

$$= \Phi_n^\mu\big(-[f_n - \Phi_n^\mu(f_n)]^2 - 2[\Phi_n^\mu(f_n) - f_n]T_n^\mu(f_n)\big)$$

$$= \Phi_n^\mu\big(-[f_n - \Phi_n^\mu(f_n)]^2 + 2f_n T_n^\mu(f_n)\big) \quad \text{[by (11)]}$$

$$= \mathrm{var}_{\Phi_n^\mu}[f_n] \times \mathrm{iact}_{K_n^\mu}[f_n]. \tag{19}$$

Here, for any probability measure $\nu \in \mathcal{P}(\mathbf{E}_n)$ and any ν-invariant Markov kernel K, we have defined the integrated autocorrelation time (IACT):

$$\mathrm{iact}_K[f_n] := 1 + 2\sum_{j=1}^\infty \frac{\mathrm{cov}_\nu[f_n, K^j(f_n)]}{\mathrm{var}_\nu[f_n]}.$$

If the MCMC kernels K_n^μ are perfectly mixing, that is if $K_n^\mu(\mathbf{x}_n, \cdot) = \Phi_n^\mu(\cdot)$ for all $\mathbf{x}_n \in \mathbf{E}_n$, then $\mathrm{iact}_{K_n^\mu}[f_n] = 1$, i.e. $\Phi_n^\mu C_n^\mu(f_n, f_n) = \mathrm{var}_{\Phi_n^\mu}[f_n]$, and the expressions for the asymptotic variances in Proposition 2 (as specified through (15) and (16)) simplify to those obtained in [9], [10], [20] for conventional SMC algorithms. Thus, by the decomposition from (19), the terms appearing in the asymptotic variance of the MCMC-PF are equal to those appearing in the asymptotic variance of standard PFs multiplied by the IACT associated with the MCMC kernels used to generate the particles.

To derive a variance ordering between standard PFs and MCMC-PFs, we assume in the remainder of this section that the MCMC operators are self-adjoint and *positive*, in the sense that $[\Phi_n^\mu \otimes K_n^\mu](f_n \otimes f_n) \geq 0$ for any Φ_n^μ-square integrable real-valued functions f_n. For positive (and self-adjoint) MCMC operators, the IACT terms are greater than 1 for any $f_n \in \mathcal{B}(\mathbf{E}_n)$ ([24, p. 20] as cited in [27, Theorem 3.7.1]). They thus represent the variance 'penalty' incurred due to the additional between-particle positive correlations in MCMC-PFs relative to standard PFs. We note that this assumption is mild since the MCMC kernels which could be deployed in practical settings are almost always positive. Examples of positive operators include the independent MH kernel [25] discussed in Example 1, the MH kernel with Gaussian or Student-t random walk proposals [4] or autoregressive positively correlated proposals with normal or Student-t innovations [13], and some versions of the hit-and-run and slice sampling algorithms [30].

4.2. Variance–variance trade-off

There is an efficiency trade-off involved in deciding whether to employ a standard PF or an MCMC-PF for a particular application. For the same distribution flow $(\eta_n)_{n\geq1}$, the former always has a lower asymptotic variance than the latter if the MCMC draws are positively correlated. However, as we seek to illustrate in the remainder of this section, an MCMC-PF may still be preferable (in terms of asymptotic variance) to a standard PF in certain situations, even if a positive MCMC kernel is used, because it can sometimes be used to target a more efficient distribution flow, i.e. a flow for which the variance terms $\mathrm{var}_{\Phi_n^\mu}[f_n]$ are reduced far enough to compensate for the IACT-based 'penalty' terms $\mathrm{iact}_{K_n^\mu}[f_n]$ in (19). Additionally, the computational cost of generating one particle in an MCMC-PF can be smaller than the corresponding cost in a standard PF.

As an illustration, we compare the asymptotic variances of approximations π_n^N of the filter π_n computed using either the standard PFs or MCMC-PFs targeting the BPF and FA-APF flows in the state-space model from Subsection 2.5. In the remainder of this section, we let $S_{p,n} \colon \mathcal{B}(\mathbf{E}_p) \to \mathcal{B}(\mathbf{E}_n)$ be a kernel that satisfies $\pi_p S_{p,n} = \pi_n$ and is given by

$$S_{p,n}(f_n)(\mathbf{x}_p) := \frac{\mathcal{L}_p}{\mathcal{L}_n} \int_{\mathbf{E}_n} f_n(\mathbf{z}_n) \delta_{\mathbf{x}_p}(\mathrm{d}\mathbf{z}_p) \prod_{q=p+1}^{n} g_q(z_q, y_q) L_q(z_{q-1}, \mathrm{d}z_q).$$

We begin by deriving expressions for the asymptotic variances in each case.

- **BPF flow.** For the BPF flow from Example 2, expectations under the filter $\pi_n(f_n) = \eta_n(G_n f_n)/\eta_n(G_n)$ are approximated by $\pi_n^N(f_n) = \eta_n^N(G_n f_n)/\eta_n^N(G_n)$. Accounting for this transformation, e.g. as in [17], yields

$$[\pi_n^N - \pi_n](f_n) = \frac{\eta_n(G_n)}{\eta_n^N(G_n)} \frac{\eta_n^N(G_n[f_n - \pi_n(f_n)])}{\eta_n(G_n)}$$

$$= \frac{\eta_n(G_n)}{\eta_n^N(G_n)} [\eta_n^N - \eta_n](G_n[f_n - \pi_n(f_n)]/\eta_n(G_n)).$$

As Corollary 1 ensures that $\eta_n(G_n)/\eta_n^N(G_n) \to_{\text{a.s.}} 1$, Slutsky's lemma and Proposition 2 are sufficient to show that for the BPF and MCMC-BPF, respectively, $\sqrt{N}[\pi_n^N - \pi_n](f_n)$ converges in distribution to a Gaussian distribution with zero mean and variance

$$\Sigma_n^{\text{BPF}}(f_n) = \sum_{p=1}^{n} \text{var}_{\tilde{\pi}_p}[\tilde{f}_{p,n}], \tag{20}$$

$$\Sigma_n^{\text{MCMC-BPF}}(f_n) = \sum_{p=1}^{n} \text{var}_{\tilde{\pi}_p}[\tilde{f}_{p,n}] \times \text{iact}_{K_p^{\tilde{\pi}_{p-1}}}[\tilde{f}_{p,n}], \tag{21}$$

where, using that $\eta_n(G_n[f_n - \pi_n(f_n)]/\eta_n(G_n)) = 0$,

$$\tilde{f}_{p,n}(\mathbf{x}_p) := \overline{Q}_{p,n}(G_n[f_n - \pi_n(f_n)]/\eta_n(G_n))(\mathbf{x}_p)$$

$$= g_p(x_p, y_p) \frac{\mathcal{L}_{p-1}}{\mathcal{L}_p} S_{p,n}(f_n - \pi_n(f_n))(\mathbf{x}_p).$$

- **FA-APF flow.** For the FA-APF flow from Example 3, $\pi_n = \eta_n$, so that we may approximate the filter by $\pi_n^N := \eta_n^N$. Hence, Proposition 2 shows that for the FA-APF and MCMC-FA-APF, respectively, $\sqrt{N}[\pi_n^N - \pi_n](f_n)$ converges in distribution to a Gaussian distribution with zero mean and variance

$$\Sigma_n^{\text{FA-APF}}(f_n) = \sum_{p=1}^{n} \text{var}_{\pi_p}[f_{p,n}], \tag{22}$$

$$\Sigma_n^{\text{MCMC-FA-APF}}(f_n) = \sum_{p=1}^{n} \text{var}_{\pi_p}[f_{p,n}] \times \text{iact}_{K_p^{\pi_{p-1}}}[f_{p,n}], \tag{23}$$

where

$$f_{p,n}(\mathbf{x}_p) := \overline{Q}_{p,n}(f_n - \pi_n(f_n))(\mathbf{x}_p) = S_{p,n}(f_n - \pi_n(f_n))(\mathbf{x}_p).$$

For the remainder of this section, assume that the asymptotic variance of the standard FA-APF is lower than that of the standard BPF for the given state-space model. More precisely, we assume that for each $p \leq n$,

$$\text{var}_{\pi_p}[f_{p,n}] \leq \text{var}_{\tilde{\pi}_p}[\tilde{f}_{p,n}],$$

and hence that

$$\Sigma_n^{\text{FA-APF}}(f_n) \leq \Sigma_n^{\text{BPF}}(f_n).$$

This is thought to hold in many applications and has been empirically verified e.g. in [35], although it is possible to construct counterexamples [18]. Assuming that the MCMC kernels K_p^{μ} are positive operators, the IACTs take values in $[1, \infty)$, and hence

$$\Sigma_n^{\text{BPF}}(f_n) \leq \Sigma_n^{\text{MCMC-BPF}}(f_n) \quad \text{and} \quad \Sigma_n^{\text{FA-APF}}(f_n) \leq \Sigma_n^{\text{MCMC-FA-APF}}(f_n).$$

However, as noted in Example 3, there are many scenarios where FA-APF cannot be implemented as we cannot generate N (conditionally) IID samples from Φ_n^N. In this case, practitioners typically have to resort to using the standard BPF instead. In contrast, the MCMC-FA-APF can usually be implemented. In such circumstances, use of MCMC-PFs (specifically in the form of the MCMC-FA-APF) can be preferable, e.g. if the variance reductions attained by targeting the FA-APF flow are large enough to outweigh the additional variance due to the increased particle correlation, i.e. if for each $1 \leq p \leq n$,

$$\text{var}_{\pi_p}[f_{p,n}] \times \text{iact}_{K_p^{\pi_{p-1}}}[f_{p,n}] \leq \text{var}_{\tilde{\pi}_p}[\tilde{f}_{p,n}],$$

because then

$$\Sigma_n^{\text{MCMC-FA-APF}}(f_n) \leq \Sigma_n^{\text{BPF}}(f_n).$$

4.3. Numerical illustration

We end this section by illustrating the 'variance–variance trade-off' mentioned above on two instances of the state-space model from Subsection 2.5.

The first model is a state-space model on a binary space $E = F := \{0, 1\}$ and with $n = 2$ observations: $y_1 = y_2 = 0$. Furthermore, for some $\alpha, \varepsilon \in [0, 1]$ and for any $x_1, x_2 \in E$, $\mu \in \mathcal{P}(\mathbf{E}_{n-1})$ and any $n \in \{1, 2\}$,

$$L_1(\{x_1\}) := 1/2, \quad L_2(x_1, \{x_2\}) := \alpha \mathbf{1}\{x_2 = x_1\} + (1 - \alpha)\mathbf{1}\{x_2 \neq x_1\},$$

$$g_n(x_n, y_n) := 0.99 \mathbf{1}\{y_n = x_n\} + 0.01 \mathbf{1}\{y_n \neq x_n\},$$

$$K_n^{\mu}(\mathbf{x}_n, \cdot) := \varepsilon \delta_{\mathbf{x}_n} + (1 - \varepsilon)\Phi_n^{\mu}.$$

While this is clearly only a toy model, we consider it for two reasons. Firstly, it allows us to analytically evaluate the asymptotic variances for standard PFs and MCMC-PFs given in (20), (21), (22) and (23). Secondly, as discussed in [18], the model allows us to select the parameter α in such a way that the FA-APF has either a lower or a higher asymptotic variance than the BPF.

Figure 1 displays the asymptotic variances relative to the asymptotic variance of the standard BPF for the test function $f_2(\mathbf{x}_2) = x_2$ for two different values of the parameter α. As displayed in the first panel, a relatively large value of α leads to the somewhat contrived case
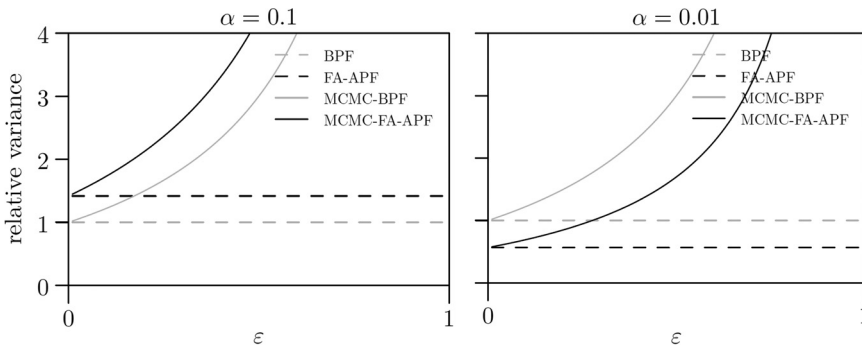
FIGURE 1: Asymptotic variances (relative to the asymptotic variance of the BPF) of the algorithms discussed in Subsection 4.2 in the case that the BPF flow is more efficient than the FA-APF flow (first panel) and in the case that the BPF flow is less efficient than the FA-APF flow (second panel).

that the BPF is more efficient than the FA-APF. However, as displayed in the second panel, a small value of α makes the FA-APF more efficient than the BPF. This is because if the system is in state 0 at time 1, the time-2 proposal used by the FA-APF incorporates the observation $y_2 = 0$, whereas the time-2 proposal used by the BPF almost always proposes moves to state 1. In this case, the MCMC-FA-APF then outperforms the BPF as long as the autocorrelation of the MCMC kernels used by the former, ε, is not too large.

We stress that in practical situations, one might expect a much more pronounced difference between the performance of the FA-APF and the BPF than is observed in this toy model; hence Markov kernels with rather modest mixing properties can sometimes still give rise to an MCMC-FA-APF that outperforms the BPF. Indeed, such algorithms showed dramatic improvements in the practical applications in [32], and the MCMC-FA-APF also outperforms the BPF in our second model, discussed below.

The second model is a d-dimensional linear Gaussian state-space model. Let $E = F := \mathbb{R}^d$ and write realisations of the d-dimensional state and observation vectors at time n as $x_n = (x_{n,i})_{1 \leq i \leq d}$ and $y_n = (y_{n,i})_{1 \leq i \leq d}$, respectively; then

$$\frac{\mathrm{d}L_1}{\mathrm{d}\lambda^{\otimes d}}(x_1) = \prod_{i=1}^{d} \phi(x_{1,i}), \qquad \frac{\mathrm{d}L_n(x_{n-1}, \cdot)}{\mathrm{d}\lambda^{\otimes d}}(x_n) = \prod_{i=1}^{d} \phi(x_{n,i} - x_{n-1,i}/2),$$

$$g_n(x_n, y_n) = \prod_{i=1}^{d} \phi(x_{n,i} - y_{n,i}),$$

where λ denotes the Lebesgue measure on $\mathbb{R}$ and ϕ denotes a Lebesgue-density of a univariate standard normal distribution. We take $K_n^\mu(\mathbf{x}_n, \cdot)$ to be an MH kernel with proposal

$$\Pi_n^\mu(\mathbf{x}_n, \mathrm{d}\mathbf{z}_n) = \begin{cases} R(x_1, \mathrm{d}z_1) & \text{if } n = 1, \\ \dfrac{F_{n-1}(\mathbf{z}_{n-1})}{\mu(F_{n-1})} \mu(\mathrm{d}\mathbf{z}_{n-1}) R(x_n, \mathrm{d}z_n) & \text{if } n > 1, \end{cases}$$

where R is a Gaussian random-walk kernel on E with transition density

$$\frac{\mathrm{d}R(x_n, \cdot)}{\mathrm{d}\lambda^{\otimes d}}(z_n) = \prod_{i=1}^{d} \sqrt{d}\phi(\sqrt{d}[z_{n,i} - x_{n,i}]),$$
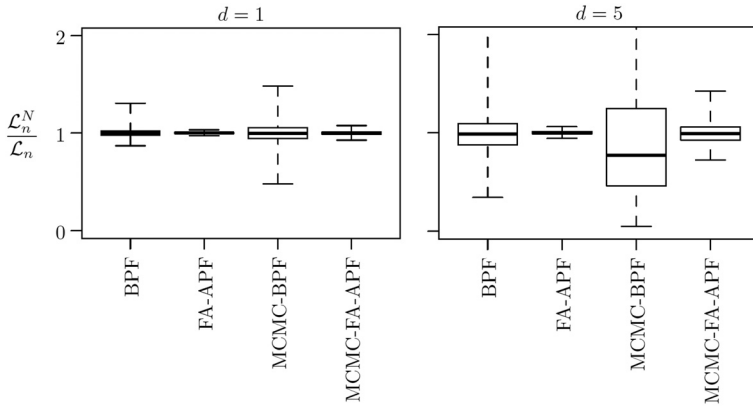
FIGURE 2: Relative estimates of the marginal likelihood $\mathcal{L}_n$ in the linear Gaussian state-space model, generated by the algorithms discussed in Subsection 4.2 using $N = 10,000$ particles (with the MCMC-FA-APF using $N = 10,000 - N_{\text{burnin}}$ particles to compensate for the additional cost of generating the samples discarded as burn-in). Based on 1,000 independent runs of each algorithm, each run using a different observation sequence of length $n = 10$ sampled from the model. For the BPF flow, $\mathcal{L}_n = \mathcal{Z}_{n+1} = \gamma_n(G_n)$ is estimated by $\mathcal{L}_n^N := \gamma_n^N(G_n)$; for the FA-APF flow, $\mathcal{L}_n = \mathcal{Z}_n = \gamma_n(\mathbf{1})$ is estimated by $\mathcal{L}_n^N := \mathcal{Z}_n^N = \gamma_n^N(\mathbf{1})$.

and where $F_n(\mathbf{x}_n) = g_n(x_n, y_n)$ for the MCMC-BPF and $F_n \equiv 1$ for the MCMC-FA-APF.

For the MCMC-BPF, the MCMC chains at each time step are initialised from stationarity, i.e.

$$\kappa_n^\mu(\mathrm{d}\mathbf{x}_n) = \Phi_n^\mu(\mathrm{d}\mathbf{x}_n) = \frac{\mu(\mathrm{d}\mathbf{x}_{n-1})g_{n-1}(x_{n-1}, y_{n-1})}{\mu(g_{n-1}(\,\cdot\,, y_{n-1}))} L_n(x_{n-1}, \mathrm{d}x_n),$$

as this is almost always possible in practice. For the MCMC-FA-APF, the MCMC chains are initialised by discarding the first $N_{\text{burnin}} = 100$ samples as burn-in, i.e. $\kappa_n^\mu = [\mu \otimes L_n](K_n^\mu)^{N_{\text{burnin}}}$.

Figure 2 displays estimates of the marginal likelihood relative to the true marginal likelihood obtained from the (MCMC-)BPF and (MCMC-)FA-APF. In this case, the MCMC-FA-APF outperforms the BPF in both dimension $d = 1$ and $d = 5$.

Note that Assumptions **A1**–**A2** and **B1**–**B3** are violated in this example. The results therefore appear to lend some support to the conjecture that these assumptions are stronger than necessary for the results of Propositions 1–2 to hold.

5. Conclusion

In this work, we have established an $\mathbb{L}_r$-inequality (which implies an SLLN) and a central limit theorem for a class of algorithms known as sequential MCMC methods or or MCMC particle filters (MCMC-PFs) and provided conditions under which the associated errors can be controlled uniformly in time. When the MCMC-PFs are based around MCMC operators that are positive (in the sense of inducing positive autocorrelations of all orders), the asymptotic variances of particle filter (PF) estimators are always lower than those of the corresponding MCMC-PF estimators. However, even if the MCMC kernels provide positively correlated draws, MCMC-PFs can remain of practical interest compared to PFs. Indeed, there are many scenarios in which a sophisticated PF such as the fully-adapted auxiliary particle

filter (FA-APF) would significantly outperform a bootstrap particle filter (BPF) but cannot be implemented, whereas the corresponding MCMC-FA-APF is essentially always applicable. If the MCMC operators used within the MCMC-FA-APF are thus displaying 'reasonable' integrated autocorrelation time, the asymptotic variance of the resulting estimators can be smaller than that of an implementable PF such as the BPF.

Appendix A. Sampling error decomposition

A.1. Proof strategy

Recall that given an approximation η_{p-1}^N of η_{p-1}, Algorithm 1 replaces $\Phi_p^{\eta_{p-1}^N}$ by a sampling-approximation η_p^N which is generated via an MCMC algorithm. This strategy is necessary because Φ_p^μ is typically intractable for any probability measure μ. However, it introduces an additional *local error* at each time step.

In this section, we relate the 'global' approximation error at time n, $\eta_n^N(f_n) - \eta_n(f_n)$, to the local errors introduced at times $1 \leq p \leq n$ via a well-known telescoping-sum decomposition. Key to the proofs of our main results is then a martingale approximation of the local errors given in Section B. The proofs of Propositions 1 and 2 and are then given in Section C. Specifically, in Subsection C.2, we give a proof of the $\mathbb{L}_r$-inequality (Proposition 1) which relies on induction (in the time index p). In this context, the martingale approximation allows us to appeal to the Burkholder–Davis–Gundy inequality for martingales to control the local error at time p in $\mathbb{L}_r$. Similarly, in Subsection C.3, we give a proof of the CLT (Proposition 2). In this context, the martingale approximation allows us to appeal to the CLT for triangular arrays of martingale difference sequences in order to show that the local errors converge in distribution to independent centred Gaussian random variables.

To simplify the notation, for any $n \geq 1$, we will hereafter often write

$$\Phi_n^N := \Phi_n^{\eta_{n-1}^N}, \quad K_n^N := K_n^{\eta_{n-1}^N}, \quad T_n^N := T_n^{\eta_{n-1}^N}, \quad C_n^N := C_n^{\eta_{n-1}^N},$$
$$\Phi_n := \Phi_n^{\eta_{n-1}}(=\eta_n), \quad K_n := K_n^{\eta_{n-1}}, \quad T_n := T_n^{\eta_{n-1}}, \quad C_n := C_n^{\eta_{n-1}}.$$

In addition, for any $1 \leq p \leq n$ and any $f_n \in \mathcal{B}(\mathbf{E}_n)$, we define the family of probability measures $\Phi_{p,n}^\mu$ associated with the Feynman–Kac model by $\Phi_{p,n}^\mu(f_n) := \mu Q_{p,n}(f_n)/\mu Q_{p,n}(\mathbf{1})$. Furthermore, we allow $f := (f_n)_{n \geq 1}$ to denote a sequence of test functions where for any $n \geq 1$, $f_n = (f_n^u)_{1 \leq u \leq d} \in \mathcal{B}(\mathbf{E}_n)^d$. For any $1 \leq u \leq d$, we also sometimes write $f^u := (f_n^u)_{n \geq 1}$.

A.2. Global-error decomposition

To control the overall approximation error at time n, we make use of a telescoping-sum decomposition commonly used in the analysis of Feynman-Kac models (see, e.g., [10, Chapter 7]):

$$\sqrt{N}[\eta_n^N - \eta_n](f_n) = \sqrt{N} \sum_{p=1}^n \Phi_{p,n}^{\eta_p^N}(f_n) - \Phi_{p,n}^{\Phi_p^N}(f_n). \tag{24}$$

Here, we recall that Φ_1^μ does not depend on μ, so that $\Phi_1^{\eta_0^N} = \Phi_1^N = \eta_1$ without the need for actually defining η_0^N. Note that the pth term in the sum on the right-hand side can be viewed as the part of the overall approximation error at time n that can be attributed to the *local error* introduced at time p. Specifically, the local error at time p is a consequence of propagating the

particle system from time $p - 1$ to time p not by the nonlinear semigroup of the limiting model, which would yield Φ_p^N, but by forming an approximation η_p^N of Φ_p^N via MCMC sampling. That is, the local error at time p—which is propagated to time n via the probability measures $\Phi_{p,n}^\mu$ in (24)—is given by

$$V_p^N(f_p) := \sqrt{N}[\eta_p^N - \Phi_p^N](f_p). \tag{25}$$

It is well known (see Subsection C.1 for details) that the 'strong mixing' Assumption **B2**–**B3** is sufficient to guarantee stability of the probability measures $\Phi_{p,n}^\mu$, so that the influence of earlier local errors diminishes over time. This stability property therefore plays a key rôle in establishing the time-uniform bounds on the $\mathbb{L}_r$-errors and on the asymptotic variance in Propositions 1 and 2.

A.3. Local-error decomposition

For the purpose of isolating the part of the error that is due to to a potential non-stationary initialisation of the MCMC chains, let $(\tilde{\boldsymbol{\xi}}_p^i)_{i \geq 1}$ be a Markov chain which evolves according to the same transition kernels as $(\boldsymbol{\xi}_p^i)_{i \geq 1}$ but which is initialised from stationarity, i.e. $\tilde{\boldsymbol{\xi}}_p^1 \sim \Phi_p^N$ and $\tilde{\boldsymbol{\xi}}_p^i \sim K_n^N(\tilde{\boldsymbol{\xi}}_p^{i-1}, \cdot)$, for $2 \leq i \leq N$. Furthermore, let $\tilde{\eta}_p^N := \frac{1}{N} \sum_{i=1}^N \delta_{\tilde{\boldsymbol{\xi}}_p^i}$ denote the associated occupation measure. The local error defined in (25) can then be decomposed as

$$V_p^N(f_p) = [\widetilde{V}_p^N + R_p^N](f_p),$$

where

$$\begin{aligned} \widetilde{V}_p^N(f_p) &:= \sqrt{N}[\tilde{\eta}_p^N - \Phi_p^N](f_p), \\ R_p^N(f_p) &:= \sqrt{N}[\eta_p^N - \tilde{\eta}_p^N](f_p), \end{aligned} \tag{26}$$

represent, respectively, the local error introduced at time p if the MCMC chain used to generate the sampling-approximation of Φ_p^N is initialised from stationarity and the additional error due to non-stationary initialisation. Tighter control of the latter could be obtained by explicitly coupling $(\tilde{\boldsymbol{\xi}}_p^i)_{i \geq 1}$ and $(\boldsymbol{\xi}_p^i)_{i \geq 1}$, but it is sufficient for our purposes to treat the two systems as being entirely independent.

Using the tower property of conditional expectation, it can be easily checked that $\mathbb{E}[\widetilde{V}_p^N(f_p)] = 0$ as in standard PFs. However, contrary to standard PFs, the particles $\tilde{\boldsymbol{\xi}}_p^i$ and $\tilde{\boldsymbol{\xi}}_p^j$, for $i \neq j$, are not necessarily conditionally independent given $\mathcal{F}_{p-1}^{N,N}$, where $\mathcal{F}_0^{N,N} := \{\emptyset, \Omega\}$ and, for any $p \geq 1$ and $1 \leq k \leq N$,

$$\mathcal{F}_p^{N,k} := \mathcal{F}_{p-1}^{N,N} \vee \sigma(\boldsymbol{\xi}_p^i, \tilde{\boldsymbol{\xi}}_p^i \mid 1 \leq i \leq k). \tag{27}$$

Because of the lack of conditional independence, we obtain, for any $1 \leq u \leq d$,

$$\begin{aligned} \mathbb{E}[\widetilde{V}_p^N(f_p^u)^2] &= \mathbb{E}\Big[\mathbb{E}\big(\widetilde{V}_p^N(f_p^u)^2 \big| \mathcal{F}_{p-1}^{N,N}\big)\Big] \\ &= \mathbb{E}\Big[\Phi_p^N\big[(f_p^u - \Phi_p^N(f_p^u))^2\big]\Big] \\ &\quad + 2 \sum_{j=1}^{N-1} \Big(1 - \frac{j}{N}\Big) \mathbb{E}\Big[\Phi_p^N\big[(f_p^u - \Phi_p^N(f_p^u))(K_p^N)^j(f_p^u - \Phi_p^N(f_p^u))\big]\Big]. \end{aligned} \tag{28}$$

To see this, note that (conditional on $\mathcal{F}_{p-1}^{N,N}$) the inner expectation in (28) is simply the variance of $N^{-1/2} \sum_{i=1}^{N} h(\mathbf{Y}^i)$, where $h(\mathbf{y}) := f_p^u(\mathbf{y}) - \Phi_p^N(f_p^u)$ and where $(\mathbf{Y}^i)_{i \geq 1}$ is a stationary Markov chain with invariant distribution Φ_p^N and transition kernels K_p^N. The last line then follows by exploiting stationarity of that Markov chain and grouping equivalent terms. This is a generalisation, when K_p^{μ} is not perfectly mixing, of the result for a standard PF in which $K_p^{\mu}(\mathbf{x}_p, \cdot) = \Phi_p^{\mu} = \kappa_p^{\mu}$ for all $\mathbf{x}_p \in \mathbf{E}_p$, in which case

$$\mathbb{E}[\widetilde{V}_p^N(f_p^u)^2] = \mathbb{E}[V_p^N(f_p^u)^2] = \mathbb{E}[\Phi_p^N([f_p^u - \Phi_p^N(f_p^u)]^2)].$$

Following [11], [5], we further decompose (26) as

$$\begin{aligned}
\widetilde{V}_p^N(f_p) &= \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \left[f_p(\tilde{\boldsymbol{\xi}}_p^i) - \Phi_p^N(f_p) \right] \\
&= \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \left[T_p^N(f_p)(\tilde{\boldsymbol{\xi}}_p^i) - K_p^N T_p^N(f_p)(\tilde{\boldsymbol{\xi}}_p^i) \right] \quad \text{[by (10)]} \\
&= U_p^N(f_p) + L_p^N(f_p),
\end{aligned}$$

where, letting $\tilde{\boldsymbol{\xi}}_p^{N+1}$ be a random variable distributed, independently conditional upon $\mathcal{F}_p^{N,N}$, according to $K_p^N(\tilde{\boldsymbol{\xi}}_p^N, \cdot)$, we have (weakly) defined

$$U_p^N(f_p) := \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \left[T_p^N(f_p)(\tilde{\boldsymbol{\xi}}_p^{i+1}) - K_p^N T_p^N(f_p)(\tilde{\boldsymbol{\xi}}_p^i) \right],$$

$$L_p^N(f_p) := \frac{1}{\sqrt{N}} \left[T_p^N(f_p)(\tilde{\boldsymbol{\xi}}_p^1) - T_p^N(f_p)(\tilde{\boldsymbol{\xi}}_p^{N+1}) \right].$$

This allows us to write the local error at time p as

$$V_p^N = U_p^N + L_p^N + R_p^N. \tag{29}$$

In the next section, we show that the terms L_p^N and R_p^N vanish as $N \to \infty$, so that the local sampling error at time p, V_p^N, can be approximated by U_p^N. The latter has a martingale property which is at the centre of our proofs of the $\mathbb{L}_r$-inequality from Proposition 1 and the CLT from Proposition 2.

Appendix B. Martingale construction

B.1. Martingale approximation at time p

For a given p, $1 \leq p \leq n$, and every $N \in \mathbb{N}$, define the filtration $\mathcal{F}_p^N := (\mathcal{F}_p^{N,i})_{0 \leq i \leq N}$, where $\mathcal{F}_p^{N,i}$ is given by (27) with the convention that $\mathcal{F}_p^{N,0} = \mathcal{F}_{p-1}^{N,N}$. We now construct a martingale approximation of the local error at time p as the number of particles grows. That is, we show that for $1 \leq p \leq n$ and each $N \in \mathbb{N}$, $U_p^N(f_p)$ is the terminal value of an $\mathcal{F}_p^N$-martingale (and for fixed p these martingales, indexed by N, form a triangular array), while $L_p^N(f_p)$ and $R_p^N(f_p)$ are remainder terms which vanish almost surely as $N \to \infty$.

- **Martingale.** For each $N \geq 1$, $U_p^N(f_p) = \sum_{i=1}^{N} \Delta U_p^{N,i+1}(f_p)$ is the terminal value of a martingale defined through the $\mathcal{F}_p^N$-martingale difference sequence $(\Delta U_p^{N,i+1}(f_p))_{1 \leq i \leq N}$, where

$$\Delta U_p^{N,i+1}(f_p) := \frac{1}{\sqrt{N}}\big[T_p^N(f_p)(\tilde{\xi}_p^{i+1}) - K_p^N T_p^N(f_p)(\tilde{\xi}_p^i)\big].$$

Indeed, for any $1 \le i \le N$ and any $1 \le u, v \le d$,

$$\mathbb{E}\big[\Delta U_p^{N,i+1}(f_p^u)\big|\mathcal{F}_p^{N,i}\big] = 0, \tag{30}$$

$$\mathbb{E}\big[\Delta U_p^{N,i+1}(f_p^u)\Delta U_p^{N,i+1}(f_p^v)\big|\mathcal{F}_p^{N,i}\big] = \frac{1}{N}C_p^N(f_p^u, f_p^v)(\tilde{\xi}_p^i), \tag{31}$$

where the second line follows directly from the definition in (13).

- **Remainder.** By (12), for any $1 \le u \le d$, the remainder signed measure L_p^N is bounded as

$$|L_p^N(f_p^u)| \le \frac{2}{\sqrt{N}}|T_p^N(f_p^u)| \le \frac{2\overline{T}_p\|f_p^u\|}{\sqrt{N}}. \tag{32}$$

By (8), under Assumption **A1**, for any $N, r \in \mathbb{N}$ and any $1 \le u \le d$, we have

$$\mathbb{E}\big[|R_p^N(f_p^u)|^r\big]^{\frac{1}{r}} \le \frac{\overline{T}_p\|f_p^u\|}{\sqrt{N}}. \tag{33}$$

Note that by Markov's inequality and the Borel–Cantelli lemma, (33) implies that $R_p^N(f_p^u) \to_{\text{a.s.}} 0$ as $N \to \infty$.

B.2. Martingale approximation up to time n

For any $N \ge 1$, the sum of all local errors up time n—of which we again construct a martingale approximation—is given by

$$\mathcal{V}_n^N(f) := \sum_{p=1}^n V_p^N(f_p) = \mathcal{U}_n^N(f) + \mathcal{L}_n^N(f) + \mathcal{R}_n^N(f).$$

The three quantities appearing on the right-hand side are defined as follows.

- **Martingale.** Let $\mathcal{F}^N := (\mathcal{F}_n^N)_{n\ge1}$ with $\mathcal{F}_p^N := \mathcal{F}_p^{N,N}$; then the terms

$$\mathcal{U}_n^N(f) := \sum_{p=1}^n U_p^N(f_p)$$

define an $\mathcal{F}^N$-martingale $(\mathcal{U}_n^N(f))_{n\ge1}$. As detailed in Section C.3, these martingales can be constructed by combining martingale increments from the *local* error decompositions in an appropriate lexicographic order. Again, we will treat this collection, indexed by N, as a triangular array.

- **Remainder.** Again,

$$\mathcal{L}_n^N(f) := \sum_{p=1}^n L_p^N(f_p) \quad \text{and} \quad \mathcal{R}_n^N(f) := \sum_{p=1}^n R_p^N(f_p)$$

constitute remainder terms. Note that for any $n \ge 1$ and any $1 \le u \le d$, by (32) and (33),

$$\lim_{N\to\infty} |\mathcal{L}_n^N(f^u)| = 0 \quad \text{and} \quad |\mathcal{R}_n^N(f^u)| \to_{\text{a.s.}} 0.$$

Appendix C. Convergence proofs

C.1. Auxiliary results needed for time-uniform bounds

For any $1 \leq p \leq n$ and any $f_n \in \mathcal{B}(\mathbf{E}_n)$, we define

$$r_{p,n} := \sup_{\mathbf{x}_p, \mathbf{z}_p \in \mathbf{E}_p} \frac{Q_{p,n}(\mathbf{1})(\mathbf{x}_p)}{Q_{p,n}(\mathbf{1})(\mathbf{z}_p)}, \quad P_{p,n}(f_n) := \frac{Q_{p,n}(f_n)}{Q_{p,n}(\mathbf{1})} \leq 1.$$

In the remainder of this work, whenever we restrict our analysis to test functions $f_n \in \mathcal{B}^\star(\mathbf{E}_n)^d$, we can replace $\beta(P_{p,n})$ by

$$\beta^\star(P_{p,n}) := \frac{1}{2} \sup |P_{p,n}(f_n')(\mathbf{x}_p) - P_{p,n}(f_n')(\mathbf{z}_p)|,$$

where the supremum is over all $\mathbf{x}_p, \mathbf{z}_p \in \mathbf{E}_p$ and all $f_n' \in \mathcal{B}^\star(\mathbf{E}_n)$ such that $\|f_n'\| \leq 1$.

Lemma 1. *Under Assumptions **A1** and **B1–B3**, for any $1 \leq p \leq n$,*

$$\overline{T}_p \leq \overline{T}, \quad \text{where} \quad \overline{T} := 2\bar{\iota}/\varepsilon(K) < \infty;$$
$$r_{p,n} \leq \bar{r}, \quad \text{where} \quad \bar{r} := \varepsilon(G)^{-(m+l)}\varepsilon(M)^{-1} < \infty;$$
$$\beta^\star(P_{p,n}) \leq \bar{\beta}^{\lfloor (n-p)/m \rfloor}, \quad \text{where} \quad \bar{\beta} := (1 - \varepsilon(G)^{m+l}\varepsilon(M)^2) < 1.$$

Proof. This follows by arguments similar to those used in the proof of [10, Proposition 4.3.3]. □

C.2. Auxiliary results needed for the $\mathbb{L}_r$-inequality

Lemma 2. *Under Assumption **A1**, for any $r \geq 1$, there exists $b_r < \infty$ such that for any $n \geq 1$, any $f_n \in \mathcal{B}(\mathbf{E}_n)$, and any $N \in \mathbb{N}$,*

$$\mathbb{E}\big[|U_n^N(f_n)|^r\big]^{\frac{1}{r}} \leq 2b_r \overline{T}_n \|f_n\|.$$

Proof. Without loss of generality, assume that $\|f_n\| \leq 1$. The quadratic variation associated with the martingale $U_n^N(f_n)$ satisfies

$$\sum_{i=1}^{N} \mathbb{E}\big(\Delta U_n^{N,i+1}(f_n)^2 \big| \mathcal{F}_n^{N,i}\big) = \tilde{\eta}_n^N C_n^N(f_n, f_n) \quad \text{[by (31)]}$$

$$\leq 4\overline{T}_n^2 \quad \text{[by (14)]}.$$

Hence, by the Burkholder–Davis–Gundy inequality (see, e.g., [19, Theorem 17.7]) there exists $b_r < \infty$ such that

$$\mathbb{E}\big[|U_n^N(f_n)|^r\big]^{\frac{1}{r}} \leq 2b_r \overline{T}_n.$$

This completes the proof. □

We are now ready to prove the $\mathbb{L}_r$-inequality in Proposition 1.

Proof. Without loss of generality, assume that $\|f_n\| \leq 1$ for all $n \geq 1$ and that the constants b_r in Lemma 2 satisfy $\inf_{r \geq 1} b_r \geq 1$.

We begin by proving the first part of the proposition, i.e. the $\mathbb{L}_r$-error bound without the additional Assumptions **B1**–**B3**. The proof proceeds by induction on n by similar arguments as in [7]. At time $n = 1$, by Minkowski's inequality combined with Lemma 2 as well as (32) and (33), we have

$$
\begin{aligned}
\sqrt{N}\mathbb{E}\big[|[\eta_1^N - \eta_1](f_1)|^r\big]^{\frac{1}{r}} = \sqrt{N}\mathbb{E}\big[|[\eta_1^N - \Phi_1^N](f_1)|^r\big]^{\frac{1}{r}} \\
\leq \mathbb{E}\big[|U_1^N(f_1)|^r\big]^{\frac{1}{r}} + \mathbb{E}\big[|L_1^N(f_1)|^r\big]^{\frac{1}{r}} + \mathbb{E}\big[|R_1^N(f_1)|^r\big]^{\frac{1}{r}} \\
\leq 2b_r\overline{T}_1 + \frac{3\overline{T}_1}{\sqrt{N}} \leq a_1 b_r,
\end{aligned}
$$

e.g. with $a_1 := 5\overline{T}_1 < \infty$.

Assume now that the first part of the proposition holds at time $n - 1$, for some $n > 1$. By Minkowski's inequality,

$$
\begin{aligned}
\sqrt{N}\mathbb{E}\big[|[\eta_n^N - \eta_n](f_n)|^r\big]^{\frac{1}{r}} \\
\leq \sqrt{N}\mathbb{E}\big[|[\eta_n^N - \Phi_n^N](f_n)|^r\big]^{\frac{1}{r}} + \sqrt{N}\mathbb{E}\big[|[\Phi_n^N - \eta_n](f_n)|^r\big]^{\frac{1}{r}}
\end{aligned}
\tag{34}
$$

$$
\leq 2b_r\overline{T}_n + \frac{3\overline{T}_n}{\sqrt{N}} + \frac{2b_r a_{n-1}}{\eta_{n-1}(G_{n-1})} \leq a_n b_r,
\tag{35}
$$

e.g. with $a_n := 5\overline{T}_n + 2a_{n-1}/\eta_{n-1}(G_{n-1}) < \infty$. Here, the bound on the first term in (34) follows by the same arguments as at time 1. The bound on the second term in (34) follows from the following decomposition (note that $\eta_{n-1}(Q_n(\mathbf{1})) = \eta_{n-1}(G_{n-1})$):

$$
\begin{aligned}
\eta_{n-1}(G_{n-1})|[\Phi_n^N - \eta_n](f_n)| \\
= \big|\eta_{n-1}(G_{n-1})\Phi_n^N(f_n) - \eta_{n-1}^N(Q_n(f_n)) + \eta_{n-1}^N(Q_n(f_n)) - \eta_{n-1}(Q_n(f_n))\big| \\
= \big|\Phi_n^N(f_n)[\eta_{n-1} - \eta_{n-1}^N](Q_n(\mathbf{1})) + [\eta_{n-1}^N - \eta_{n-1}](Q_n(f_n))\big| \\
\leq \|\Phi_n^N(f_n)\||[\eta_{n-1} - \eta_{n-1}^N](Q_n(\mathbf{1}))| + |[\eta_{n-1}^N - \eta_{n-1}](Q_n(f_n))|.
\end{aligned}
$$

Minkowski's inequality along with $\|Q_n(\mathbf{1})\| \leq 1$ and $\|Q_n(f_n)\| \leq 1$ as well as the bound $\|\Phi_n^N(f_n)\| \leq \|f_n\| \leq 1$ combined with the induction assumption then readily yields the bound given in (35), i.e.

$$
\sqrt{N}\mathbb{E}\big[|[\Phi_n^N - \eta_n](f_n)|^r\big]^{\frac{1}{r}} \leq \frac{2b_r a_{n-1}}{\eta_{n-1}(G_{n-1})}.
$$

This completes the first part of the proposition.

As the bounds obtained through the previous induction proof cannot easily be made time-uniform, we prove the second part of the proposition via the more conventional telescoping-sum decomposition given in (24). Using the arguments in [10, pp. 244–246], we obtain the following bound for the pth term in the telescoping sum:

$$
\begin{aligned}
\sqrt{N}|[\Phi_{p,n}^{\eta_p^N} - \Phi_{p,n}^{\Phi_p^N}](f_n)| \leq 2\sqrt{N}|[\eta_p^N - \Phi_p^N](\overline{Q}_{p,n}^N(f_n))|r_{p,n}\beta(P_{p,n}) \\
= 2|[U_p^N + L_p^N + R_p^N](\overline{Q}_{p,n}^N(f_n))|r_{p,n}\beta(P_{p,n}),
\end{aligned}
$$

where the second line is due to (29) and where

$$\overline{Q}_{p,n}^N(f_n) := \frac{Q_{p,n}^N(f_n)}{\|Q_{p,n}^N(f_n)\|},$$

$$Q_{p,n}^N(f_n) := \frac{Q_{p,n}(\mathbf{1})}{\Phi_p^N(Q_{p,n}(\mathbf{1}))} P_{p,n}\left(f_n - \frac{\Phi_p^N(Q_{p,n}(f_n))}{\Phi_p^N(Q_{p,n}(\mathbf{1}))}\right).$$

Hence, by (24),

$$\sqrt{N}\mathbb{E}\big[|[\eta_n^N - \eta_n](f_n)|^r\big]^{\frac{1}{r}}$$

$$\leq 2 \sum_{p=1}^{n} \mathbb{E}\big[|[U_p^N + L_p^N + R_p^N](\overline{Q}_{p,n}^N(f_n))|^r\big]^{\frac{1}{r}} r_{p,n}\beta(P_{p,n})$$

$$\leq 2 \sum_{p=1}^{n} \left(2b_r\overline{T}_p + \frac{3\overline{T}_p}{\sqrt{N}}\right) r_{p,n}\beta(P_{p,n}) \leq a_n b_r$$

with $a_n := 10 \sum_{p=1}^{n} \overline{T}_p r_{p,n}\beta(P_{p,n})$, where the last line follows from Minkowski's inequality combined with Lemma 2, (32), and (33). Since $f_n \in \mathcal{B}^\star(\mathbf{E}_n)$, we can replace $\beta(P_{p,n})$ by $\beta^\star(P_{p,n})$ in the derivation above. Lemma 1 then yields the time-uniform bound

$$a_n \leq 10\overline{T}\overline{r} \sum_{p=1}^{n} \overline{\beta}^{\lfloor(n-p)/m\rfloor} \leq 10\overline{T}\overline{r}m \sum_{n=0}^{\infty} \overline{\beta}^n \leq \frac{20\overline{\iota}m}{\varepsilon(K)\varepsilon(M)^3\varepsilon(G)^{2(m+l)}} =: a.$$

This completes the proof.

C.3. Auxiliary results needed for the CLT

In this subsection, we prove the CLT stated in Proposition 2. The proof relies on Lemma 4, which establishes convergence of the covariance function, and whose proof in turn relies on Lemma 3. The latter illustrates how local Lipschitz conditions of the type given in Assumption **A2** are used within this work.

Lemma 3. *For measurable spaces $(H, \mathcal{H})$ and $(H', \mathcal{H}')$, let $M^\nu : \mathcal{B}(H') \to \mathcal{B}(H)$ be bounded integral operators indexed by $\nu \in \mathcal{P}(H)$. Let $(\mu^N)_{N \geq 1}$ be a sequence of random probability measures on $(H, \mathcal{H})$ and $\mu \in \mathcal{P}(H)$. Assume that*

(1) *for any $r \geq 1$, there exists $b_r < \infty$ such that for any $g \in \mathcal{B}(H)$ and any $N \geq 1$,*

$$\mathbb{E}\big[|[\mu^N - \mu](g)|^r\big]^{\frac{1}{r}} \leq \frac{b_r}{\sqrt{N}}\|g\|; \tag{36}$$

(2) *there exist a constant $\overline{\Gamma} < \infty$ and a family of bounded integral operators $(\Gamma^\nu)_{\nu \in \mathcal{P}(H)}$ from $\mathcal{B}(H')$ into $\mathcal{B}(H)$ such that for any $f \in \mathcal{B}(H')$ and any $\nu_0, \nu_1 \in \mathcal{P}(H)$,*

$$\|[M^{\nu_1} - M^{\nu_0}](f)\| \leq \int_{\mathcal{B}(H)} |[\nu_1 - \nu_0](g)| \Gamma^{\nu_0}(f, \mathrm{d}g), \tag{37}$$

with

$$\int_{\mathcal{B}(H)} \|g\| \Gamma^{\nu_0}(f, \mathrm{d}g) \leq \|f\|\overline{\Gamma}. \tag{38}$$

Then for any $r \geq 1, f \in \mathcal{B}(H')$, and $N \geq 1$,

$$\mathbb{E}\big[\|[M^{\mu^N} - M^{\mu}](f)\|^r\big]^{\frac{1}{r}} \leq \frac{b_r \overline{\Gamma}}{\sqrt{N}} \|f\|.$$

Proof. For any $r \geq 1$,

$$
\begin{aligned}
\mathbb{E}[\|[M^{\mu^N} - M^{\mu}](f)\|^r]^{\frac{1}{r}} &\leq \mathbb{E}\Bigg[\Bigg|\int_{\mathcal{B}(H)} |[\mu^N - \mu](g)| \Gamma^{\mu}(f, \mathrm{d}g)\Bigg|^r\Bigg]^{\frac{1}{r}} \\
&\leq \int_{\mathcal{B}(H)} \mathbb{E}[|[\mu^N - \mu](g)|^r]^{\frac{1}{r}} \Gamma^{\mu}(f, \mathrm{d}g) \\
&\leq \frac{b_r}{\sqrt{N}} \int_{\mathcal{B}(H)} \|g\| \Gamma^{\mu}(f, \mathrm{d}g) \\
&\leq \frac{b_r \overline{\Gamma}}{\sqrt{N}} \|f\|.
\end{aligned}
$$

Here, the first line follows from (37); the second line is due to the convexity of the $\mathbb{L}_r$-norm; the last two lines follow from (36) and (38), respectively.

Lemma 4. *Under Assumptions A1–A2, for any $n \geq 1$, there exists $\overline{\Upsilon}_n < \infty$ such that for any $N, r \geq 1$ and any $(f_n, g_n) \in \mathcal{B}(\mathbf{E}_n)^2$,*

$$\mathbb{E}\big[\|C_n^{\eta_{n-1}^N}(f_n, g_n) - C_n^{\eta_{n-1}}(f_n, g_n)\|^r\big]^{\frac{1}{r}} \leq \frac{a_{n-1} b_r \overline{\Upsilon}_n}{\sqrt{N}} \|f_n\|\|g_n\|,$$

where $a_{n-1}, b_r < \infty$ are from Proposition 1.

Proof. We use a similar argument to that used in the first part of the proof of [5, Theorem 3.5]. That is, under Assumptions A1 and A2, and using [5, Proposition 3.1], a telescoping-sum decomposition allows us to find a constant $\overline{\Upsilon}_n < \infty$ and a family of bounded integral operators $(\Upsilon_n^{\nu})_{\nu \in \mathcal{P}(\mathbf{E}_{n-1})}$ from $\mathcal{B}(\mathbf{E}_n)^2$ into $\mathcal{B}(\mathbf{E}_{n-1})$ such that for any $(\nu, \mu) \in \mathcal{P}(\mathbf{E}_{n-1})^2$,

$$\|C_n^{\nu}(f_n, g_n) - C_n^{\mu}(f_n, g_n)\| \leq \int_{\mathcal{B}(\mathbf{E}_{n-1})} |[\nu - \mu](h)| \Upsilon_n^{\mu}((f_n, g_n), \mathrm{d}h),$$

with

$$\int_{\mathcal{B}(\mathbf{E}_{n-1})} \|h\| \Upsilon_n^{\mu}((f_n, g_n), \mathrm{d}h) \leq \|f_n\|\|g_n\|\overline{\Upsilon}_n.$$

The proof is then completed by appealing to Lemma 3.

We now prove the following proposition, which adapts [5, Proposition 4.3] (see also [10, Theorem 9.3.1]) to our setting.

Proposition 3. *Let $f := (f_n)_{n \geq 1}$, where $f_n = (f_n^u)_{1 \leq u \leq d} \in \mathcal{B}(\mathbf{E}_n)^d$. The sequence of martingales $\mathcal{U}^N(f) = (\mathcal{U}_n^N(f))_{n \geq 1}$ converges in law as $N \to \infty$ to a Gaussian martingale $\mathcal{U}(f) = (\mathcal{U}_n(f))_{n \geq 1}$ such that for any $n \geq 1$ and any $1 \leq u, v \leq d$,*

$$\langle \mathcal{U}(f^u), \mathcal{U}(f^v)\rangle_n = \sum_{p=1}^{n} \eta_p C_p(f_p^u, f_p^v).$$

Proof. We begin by re-indexing the processes defined above. For any $(p, i) \in \mathbb{N} \times \{1, \ldots, N\}$, define the bijection θ^N by

$$\theta^N(p, i) := (p-1)N + i - 1.$$

Define the filtration $\mathcal{G}^N := (\mathcal{G}_k^N)_{k \geq 1}$, where $\mathcal{G}_k^N := \vee_{(p,i):\, \theta^N(p,i) \leq k} \mathcal{F}_p^{N,i}$. We then have $\mathcal{U}_n^N(f) = \widetilde{\mathcal{U}}_k^N(f)$ whenever $k = \theta^N(n, N)$, where

$$\widetilde{\mathcal{U}}_k^N(f) := \sum_{j=1}^k \Delta \widetilde{\mathcal{U}}_j^N(f)$$

defines a $\mathcal{G}^N$-martingale $(\widetilde{\mathcal{U}}_k^N(f))_{k \geq 1}$ with increments

$$\Delta \widetilde{\mathcal{U}}_j^N(f) := \Delta U_p^{N,i}(f_p) \quad \text{for } \theta^N(p, i) = j.$$

Indeed, by (30) and (31), for any $1 \leq u, v \leq d$,

$$\mathbb{E}\big[\Delta \widetilde{\mathcal{U}}_j^N(f^u)\big|\mathcal{G}_{j-1}^N\big] = 0,$$

$$\mathbb{E}\big[\Delta \widetilde{\mathcal{U}}_j^N(f^u)\Delta \widetilde{\mathcal{U}}_j^N(f^v)\big|\mathcal{G}_{j-1}^N\big] = \frac{1}{N} C_p^N(f_p^u, f_p^v)(\tilde{\xi}_p^i),$$

where $p \geq 1$ and $1 \leq i \leq N$ satisfy $\theta^N(p, i) = j$.

We now apply the CLT for triangular arrays of martingale difference sequences (see, e.g., [34, Section VII.8, Theorem 4; p. 543]). The Lindeberg condition is satisfied because the test functions f_n are bounded. Finally, for any $p \geq 1$,

$$\sum_{k=(pN)+1}^{(p+1)N} \mathbb{E}[\Delta \widetilde{\mathcal{U}}_k^N(f^u)\Delta \widetilde{\mathcal{U}}_k^N(f^v)|\mathcal{G}_{k-1}^N] = \frac{1}{N}\sum_{i=1}^N C_p^N(f_p^u, f_p^v)(\tilde{\xi}_p^i)$$

$$= \tilde{\eta}_p^N C_p^N(f_p^u, f_p^v)$$

$$\to_{\text{a.s.}} \eta_p C_p(f_p^u, f_p^v).$$

The last line follows by Markov's inequality combined with the Borel–Cantelli lemma after noting that for any $r, p \geq 1$, Minkowski's inequality combined with (33), Lemma 4, and Proposition 1 guarantees the existence of $a_p', b_r' < \infty$ such that for any $N \geq 1$,

$$\mathbb{E}\big[\|\tilde{\eta}_p^N C_p^N(f_p^u, f_p^v) - \eta_p C_p(f_p^u, f_p^v)\|^r\big]^{\frac{1}{r}} \leq \mathbb{E}\big[\|[\tilde{\eta}_p^N - \eta_p^N](C_p^N(f_p^u, f_p^v))\|^r\big]^{\frac{1}{r}}$$

$$+ \mathbb{E}\big[\|C_p^N(f_p^u, f_p^v) - C_p(f_p^u, f_p^v)\|^r\big]^{\frac{1}{r}}$$

$$+ \mathbb{E}\big[\|[\eta_p^N - \eta_p](C_p(f_p^u, f_p^v))\|^r\big]^{\frac{1}{r}}$$

$$\leq \frac{a_p' b_r'}{\sqrt{N}} \|f_n^u\| \|f_n^v\|.$$

As a result, for any $n \geq 1$, as $N \to \infty$,

$$\sum_{k=1}^{(n+1)N} \mathbb{E}[\Delta \widetilde{\mathcal{U}}_k^N(f^u)\Delta \widetilde{\mathcal{U}}_k^N(f^v)|\mathcal{G}_{k-1}^N] \to_{\text{a.s.}} \sum_{p=1}^n \eta_p C_p(f_p^u, f_p^v).$$

This completes the proof.

As an immediate consequence of Proposition 3, we obtain the following corollary. Its proof is a straightforward modification of the proof of [10, Corollary 9.3.1].

Corollary 2. *As $N \to \infty$, the sequence of random fields $V^N = (V_n^N)_{n \geq 1}$ converges in law (and in the sense of convergence of finite-dimensional marginals) to the sequence of independent and centred Gaussian random fields $V = (V_n)_{n \geq 1}$ with covariance function as defined in* (15).

We are now ready to give the proof of the CLT in Proposition 2.

Proof. The proof of the CLT now follows by replacing [10, Theorem 9.3.1 & Corollary 9.3.1] in the proofs of [10, Propositions 9.4.1 & 9.4.2] with Proposition 3 and Corollary 2, respectively.

For the time-uniform bound on the asymptotic variance in (17), we note that for any $1 \leq u \leq d$,

$$\|\overline{Q}_{p,n}(f_n^u - \eta_n(f_n^u))\| = \left\| \frac{Q_{p,n}(\mathbf{1})}{\eta_p Q_{p,n}(\mathbf{1})} \left[P_{p,n}(f_n^u) - \Psi_{p,n}^{\eta_p} P_{p,n}(f_n^u) \right] \right\|$$
$$\leq r_{p,n} \|[\mathrm{Id} - \Psi_{p,n}^{\eta_p}] P_{p,n}(f_n^u)\|$$
$$\leq 2\|f_n^u\| r_{p,n} \beta^\star(P_{p,n}),$$

where we have used that $\eta_n = \Psi_{p,n}^{\eta_p} P_{p,n}$, where

$$\Psi_{p,n}^{\eta_p}(\mathrm{d}\mathbf{x}_p) := \frac{\eta_p(\mathrm{d}\mathbf{x}_p) Q_{p,n}(\mathbf{1})(\mathbf{x}_p)}{\eta_p Q_{p,n}(\mathbf{1})}.$$

Hence, by (14) and Lemma 1, the time-uniform bound on the asymptotic variance in (18) holds, e.g. with

$$c := 8\overline{T}^2 \overline{r}^2 \sum_{p=1}^{n} \bar{\beta}^{2\lfloor (n-p)/m \rfloor} \leq 8\overline{T}^2 \overline{r}^2 m \sum_{n=0}^{\infty} \bar{\beta}^n \leq \frac{32\overline{\iota}^2 m}{\varepsilon(K)^2 \varepsilon(M)^4 \varepsilon(G)^{3(m+l)}}.$$

This completes the proof.

Acknowledgement

References

[1] ANDRIEU, C., DOUCET, A. AND HOLENSTEIN, R. (2010). Particle Markov chain Monte Carlo methods. *J. R. Statist. Soc. B [Statist. Methodology]* **72**, 269–342. With discussion.

[2] ANDRIEU, C., JASRA, A., DOUCET, A. AND DEL MORAL, P. (2011). On nonlinear Markov chain Monte Carlo. *Bernoulli* **17**, 987–1014.

[3] ANDRIEU, C. AND ROBERTS, G. O. (2009). The pseudo-marginal approach for efficient Monte Carlo computations. *Ann. Statist.* **37**, 697–725.

[4] BAXENDALE, P. H. (2005). Renewal theory and computable convergence rates for geometrically ergodic Markov chains. *Ann. Appl. Prob.* **15**, 700–738.

[5] BERCU, B., DEL MORAL, P. AND DOUCET, A. (2012). Fluctuations of interacting Markov chain Monte Carlo methods. *Stoch. Proc. Appl.* **122**, 1304–1331.

[6] BERZUINI, C., BEST, N. G., GILKS, W. R. AND LARIZZA, C. (1997). Dynamic conditional independence models and Markov chain Monte Carlo methods. *J. Amer. Statist. Soc.* **92**, 1403–1412.

[7] BROCKWELL, A., DEL MORAL, P. AND DOUCET, A. (2010). Sequentially interacting Markov chain Monte Carlo methods. *Ann. Statist* **38**, 3387–3411.

[8] CARMI, A., SEPTIER, F. AND GODSILL, S. J. (2012). The Gaussian mixture MCMC particle algorithm for dynamic cluster tracking. *Automatica* **48**, 2454–2467.

[9] CHOPIN, N. (2004). Central limit theorem for sequential Monte Carlo methods and its application to Bayesian inference. *Ann. Statist.* **32**, 2385–2411.

[10] DEL MORAL, P. (2004). *Feynman-Kac Formulae: Genealogical and Interacting Particle Systems with Applications*. Springer, New York.

[11] DEL MORAL, P. AND DOUCET, A. (2010). Interacting Markov chain Monte Carlo methods for solving nonlinear measure-valued equations. *Ann. Appl. Prob.* **20**, 593–639.

[12] DOUC, R., MOULINES, E. AND OLSSON, J. (2014). Long-term stability of sequential Monte Carlo methods under verifiable conditions. *Ann. Appl. Prob.* **24**, 1767–1802.

[13] DOUCET, A., PITT, M., DELIGIANNIDIS, G. AND KOHN, R. (2015). Efficient implementation of Markov chain Monte Carlo when using an unbiased likelihood estimator. *Biometrika* **102**, 295–313.

[14] FINKE, A., DOUCET, A. AND JOHANSEN, A. M. (2016). On embedded hidden Markov models and particle Markov chain Monte Carlo methods. Preprint. Available at http://arxiv.org/abs/1610.08962.

[15] GOLIGHTLY, A. AND WILKINSON, D. J. (2006). Bayesian sequential inference for nonlinear multivariate diffusions. *Statist. Comput.* **16**, 323–338.

[16] GORDON, N. J., SALMOND, D. J. AND SMITH, A. F. M. (1993). Novel approach to nonlinear/non-Gaussian Bayesian state estimation. *IEE Proc. F, Radar and Signal Processing* 140, 107–113.

[17] JOHANSEN, A. M. AND DOUCET, A. (2007). Auxiliary variable sequential Monte Carlo methods. Technical report 07:09. University of Bristol, Statistics Groups.

[18] JOHANSEN, A. M. AND DOUCET, A. (2008). A note on auxiliary particle filters. *Statist. Prob. Lett.* **78**, 1498–1504.

[19] KALLENBERG, O. (2006). *Foundations of Modern Probability*. Springer, New York.

[20] KÜNSCH, H. R. (2005). Recursive Monte Carlo filters: algorithms and theoretical analysis. *Ann. Statist.* **33**, 1983–2021.

[21] KÜNSCH, H. R. (2013). Particle filters. *Bernoulli* **19**, 1391–1403.

[22] LI, Y. AND COATES, M. (2017). Sequential MCMC with invertible particle flow. In *2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, IEEE, pp. 3844–3848.

[23] LIN, M., CHEN, R. AND LIU, J. S. (2013). Lookahead strategies for sequential Monte Carlo. *Statist. Sci.* **28**, 69–94.

[24] LIU, J. S. (1991). Correlation structure and convergence rate of the Gibbs sampler. Doctoral Thesis, University of Chicago.

[25] LIU, J. S. (1996). Metropolized independent sampling with comparisons to rejection sampling and importance sampling. *Statist. Comput.* **6**, 113–119.

[26] MENGERSEN, K. L. AND TWEEDIE, R. L. (1996). Rates of convergence of the Hastings and Metropolis algorithms. *Ann. Statist* **24**, 101–121.

[27] MIRA, A. (1998). Ordering, slicing and splitting Monte Carlo Markov chains. Doctoral Thesis, University of Minnesota.

[28] PAL, S. AND COATES, M. (2018). Sequential MCMC with the discrete bouncy particle sampler. In *IEEE Statistical Signal Processing Workshop (SSP)*, IEEE, pp. 663–667.

[29] PITT, M. K. AND SHEPHARD, N. (1999). Filtering via simulation: Auxiliary particle filters. *J. Amer. Statist. Soc.* **94**, 590–599.

[30] RUDOLF, D. AND ULLRICH, M. (2013). Positivity of hit-and-run and related algorithms. *Electron. Commun. Prob.* **18**, 1–8.

[31] SEPTIER, F., PANG, S. K., CARMI, A. AND GODSILL, S. (2009). On MCMC-based particle methods for Bayesian filtering: application to multitarget tracking. In *3rd IEEE International Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP)*, IEEE, pp. 360–363.

[32] SEPTIER, F. AND PETERS, G. W. (2016). Langevin and Hamiltonian based sequential MCMC for efficient Bayesian filtering in high-dimensional spaces. *IEEE J. Sel. Top. Signal Processing* 10, 312–327.

[33] SHESTOPALOFF, A. Y. AND NEAL, R. M. (2018). Sampling latent states for high-dimensional non-linear state space models with the embedded HMM method. *Bayesian Anal.* **13**, 797–822.

[34] SHIRYAEV, A. N. (1995). *Probability*, 2nd edn. Springer, New York.

[35] SNYDER, C., BENGTSSON, T. AND MORZFELD, M. (2015). Performance bounds for particle filters using the optimal proposal. *Mon. Weather Rev.* **143**, 4750–4761.

[36] WHITELEY, N. (2013). Stability properties of some particle filters. *Ann. Appl. Prob.* **23**, 2500–2537.