Submitted to the Annals of Statistics

SEQUENTIALLY INTERACTING MARKOV CHAIN MONTE CARLO METHODS

By Anthony Brockwell[*], Pierre Del Moral and Arnaud Doucet[†]

Carnegie Mellon University, INRIA Bordeaux Sud-Ouest and University of British Columbia

We introduce a novel methodology for sampling from a sequence of probability distributions of increasing dimension and estimating their normalizing constants. These problems are usually addressed using Sequential Monte Carlo (SMC) methods. The alternative Sequentially Interacting Markov Chain Monte Carlo (SIMCMC) scheme proposed here works by generating interacting non-Markovian sequences which behave asymptotically like independent Metropolis-Hastings (MH) Markov chains with the desired limiting distributions. Contrary to SMC methods, this scheme allows us to iteratively improve our estimates in an MCMC-like fashion. We establish convergence of the algorithm under realistic verifiable assumptions and demonstrate its performance on several examples arising in Bayesian time series analysis.

1. Introduction. Consider a sequence of probability distributions $\{\pi_n\}_{n\in\mathbb{T}}$ where $\mathbb{T}=\{1,2,...,P\}$, which we will refer to as "target" distributions. We shall also refer to n as the time index. For ease of presentation, we shall assume here that $\pi_n(d\mathbf{x}_n)$ is defined on a measurable space $(E_n,\mathcal{F}_n)$ where $E_1=E$, $\mathcal{F}_1=\mathcal{F}$ and $E_n=E_{n-1}\times E$, $\mathcal{F}_n=\mathcal{F}_{n-1}\times\mathcal{F}$ and we denote $\mathbf{x}_n=(x_1,...,x_n)$ where $x_i\in E$ for $i=1,...,n$. Each $\pi_n(d\mathbf{x}_n)$ is assumed to admit a density $\pi_n(\mathbf{x}_n)$ with respect to a $\sigma-$finite dominating measure denoted $d\mathbf{x}_n$ and $d\mathbf{x}_n=d\mathbf{x}_{n-1}\times dx_n$. Additionally, we have

$$\pi_n(\mathbf{x}_n)=\frac{\gamma_n(\mathbf{x}_n)}{Z_n}$$

where $\gamma_n:E_n\to\mathbb{R}^+$ is known pointwise and the normalizing constant Z_n is unknown.

---

[*]Alphabetical authorship

[†]Corresponding author

AMS 2000 subject classifications: Primary 65C05, 60J05; secondary 62F15

Keywords and phrases: Markov chain Monte Carlo, normalizing constants, sequential Monte Carlo, state-space models

In a number of important applications, it is desirable to be able to sample from the sequence of distributions $\{\pi_n\}_{n\in\mathbb{T}}$ and to estimate their normalizing constants $\{Z_n\}_{n\in\mathbb{T}}$; the most popular statistical application is the class of non-linear non-Gaussian state-space models detailed in Section 4. In this context, π_n is the posterior distribution of the hidden state variables from time 1 to n given the observations from time 1 to n and Z_n is the marginal likelihood of these observations. Many other applications - including contingency tables and population genetics - are discussed in [6], [10] and [16].

A now standard approach to solve this class of problems relies on Sequential Monte Carlo (SMC) methods; see [10] and [16] for a review of the literature. In the SMC approach, the target distributions are approximated by a large number of random samples - termed particles - which are carried forward over time by using a combination of sequential importance sampling and resampling steps. These methods have become the tools of choice for sequential Bayesian inference but, even when there is no requirement for 'real-time' inference, SMC algorithms are increasingly used as an alternative to MCMC; see for example [5], [7] and [16] for applications to econometrics models, finite mixture models and contingency tables. They also allow us to implement easily goodness-of-fit tests in a time series context -e.g. [4]- whereas a standard MCMC implementation is cumbersome [11]. Moreover, they provide an estimate of the marginal likelihood of the data.

The SMC methodology is now well-established and many theoretical convergence results are available [6]. Nevertheless, in practice, it is typically impossible to determine beforehand the number of particles necessary to achieve a fixed precision for a given application and users typically perform multiple runs for an increasing number of particles until stabilization of the Monte Carlo estimates is observed. Moreover, SMC algorithms are substantially different from MCMC algorithms and can appear difficult to implement for non-specialists.

In this paper we propose an alternative to SMC named *Sequentially Interacting Markov Chain Monte Carlo* (SIMCMC). SIMCMC methods allow us to compute Monte Carlo estimates of the quantities of interest iteratively as they are, for instance, when using MCMC methods. This allows us to refine the Monte Carlo estimates until a suitably chosen stopping time. Furthermore, for people familiar with MCMC methods, SIMCMC methods are somewhat simpler than SMC methods to implement, because they only rely on MH steps. However, SIMCMC methods are not a class of MCMC methods. These are non-Markovian algorithms which can be interpreted as an approximation of P 'ideal' standard MCMC chains. It is based on the same key idea as SMC methods; that is as $\pi_{n+1}(\mathbf{x}_n) = \int \pi_{n+1}(\mathbf{x}_n)\, dx_n$ is often 'close' to $\pi_n(\mathbf{x}_n)$, it is sensible to use $\pi_n(\mathbf{x}_n)$ as part of a proposal distribution to sample $\pi_{n+1}(\mathbf{x}_{n+1})$. In SMC methods, the correction between the proposal distribution and the target distribution is performed using Importance Sampling whereas in SIMCMC methods it is performed using an MH step. Such a strategy is computationally much more efficient than sampling separately from each target distribution using standard MCMC methods and also provides direct estimates of the normalizing constants $\{Z_n\}_{n\in\mathbb{T}}$.

The potential real-time applications are also worth commenting on. SMC methods have been used in various real-time engineering applications, for example, in neural decoding [2] and in target tracking. In these problems, it is important to be able to compute functionals of the posterior distributions of some quantity of interest, but it must also be done in real-time. SMC methods work with collections of particles that are updated sequentially to reflect these distributions. Clearly, in such real-time problems it is important that the collections of particles are not too large, or else the computational burden can cause the SMC algorithm to fall behind the system being analyzed. SIMCMC methods provide a very convenient way to make optimal use of what computing power is available. Since SIMCMC works by adding one particle at a time to collections representing distributions, we can simply run it continually in between arrival of successive observations, and it will accrue as many particles as it can in whatever amount of time is taken.

The rest of the paper is organized as follows. In Section 2, we describe SIMCMC methods, give some guidelines for the design of efficient algorithms and discuss implementation issues. In Section 3, we present some convergence results. In Section 4, we demonstrate the performance of this algorithm for various Bayesian time series problems and compare it to SMC. Finally we discuss a number of further potential extensions in Section 5. The proofs of the results in Section 3 can be found in Appendix A.

2. Sequentially Interacting Markov Chain Monte Carlo.

2.1. *The SIMCMC Algorithm.* The SIMCMC algorithm constructs P sequences $\left\{\mathbf{X}_1^{(i)}\right\}, \left\{\mathbf{X}_2^{(i)}\right\}, \ldots, \left\{\mathbf{X}_P^{(i)}\right\}$, with the property that as the iteration index i approaches infinity, the distribution of $\mathbf{X}_n^{(i)}$ approaches π_n; see Section 3. To specify the algorithm, we require a sequence of P proposal distributions, specified by their densities

$$q_1(x_1), q_2(\mathbf{x}_1, x_2), \ldots, q_P(\mathbf{x}_{P-1}, x_P).$$

Each q_n is a density in its last argument x_n with respect to dx_n, which may depend (for $n = 2, \ldots, P$) on the first argument. Proposals are drawn from $q_1(\cdot)$ for updates of the sequence $\{\mathbf{X}_1^{(i)}\}$, from $q_2(\cdot)$ for updates of the sequence $\{\mathbf{X}_2^{(i)}\}$, and so on. (Selection of proposal distributions is discussed below.) Based on these proposals, we define the weights

$$\begin{aligned} w_1(\mathbf{x}_1) &= \frac{\gamma_1(\mathbf{x}_1)}{q_1(\mathbf{x}_1)}, \\ w_n(\mathbf{x}_n) &= \frac{\gamma_n(\mathbf{x}_n)}{\gamma_{n-1}(\mathbf{x}_{n-1})\, q_n(\mathbf{x}_{n-1}, x_n)}, \quad n = 2, \ldots, P. \end{aligned} \tag{2.1}$$

For any measure μ_{n-1} on $(E_{n-1}, \mathcal{F}_{n-1})$, we define

$$(\mu_{n-1} \times q_n)(d\mathbf{x}_n) = \mu_{n-1}(d\mathbf{x}_{n-1})\, q_n(\mathbf{x}_{n-1}, dx_n)$$

and

$$\mathcal{S}_n = \{\mathbf{x}_n \in E_n : \pi_n(\mathbf{x}_n) > 0\}. \tag{2.2}$$

Intuitively, the SIMCMC algorithm proceeds as follows. At each iteration i of the algorithm, the algorithm samples $\mathbf{X}_n^{(i)}$ for $n \in \mathbb{T}$ by first sampling $\mathbf{X}_1^{(i)}$, then $\mathbf{X}_2^{(i)}$ and so on. For $n = 1$, $\left\{\mathbf{X}_1^{(i)}\right\}$ is a standard Markov chain generated using an independent MH sampler of invariant distribution π_1 and proposal distribution q_1. For $n = 2$, we would like to approximate an independent MH sampler of invariant distribution $\pi_2(\mathbf{x}_2)$ and proposal distribution $(\pi_1 \times q_2)(\mathbf{x}_2)$. As it is impossible to sample from π_1 exactly, we replace π_1 at iteration i by its current empirical measure approximation $\widehat{\pi}_1^{(i)}$. Similarly for $n > 2$, we approximate an MH sampler of invariant distribution $\pi_n(\mathbf{x}_n)$ and proposal distribution $(\pi_{n-1} \times q_n)(\mathbf{x}_n)$ by replacing π_{n-1} at iteration i by its current empirical measure approximation $\widehat{\pi}_{n-1}^{(i)}$. The sequences $\left\{\mathbf{X}_2^{(i)}\right\}, \ldots, \left\{\mathbf{X}_P^{(i)}\right\}$ generated this way are clearly non-Markovian.

Sequentially Interacting Markov Chain Monte Carlo

- Initialization, $i = 0$
  - For $n \in \mathbb{T}$, set randomly $\mathbf{X}_n^{(0)} = \mathbf{x}_n^{(0)} \in \mathcal{S}_n$.
- For iteration $i \geq 1$
  - For $n = 1$
    - Sample $\mathbf{X}_1^{*(i)} \sim q_1(\cdot)$.
    - With probability

$$\alpha_1(\mathbf{X}_1^{(i-1)}, \mathbf{X}_1^{*(i)}) = 1 \wedge \frac{w_1\left(\mathbf{X}_1^{*(i)}\right)}{w_1\left(\mathbf{X}_1^{(i-1)}\right)} \tag{2.3}$$

set $\mathbf{X}_1^{(i)} = \mathbf{X}_1^{*(i)}$, otherwise set $\mathbf{X}_1^{(i)} = \mathbf{X}_1^{(i-1)}$.

- For $n = 2, \ldots, P$
  - Sample $\mathbf{X}_n^{*(i)} \sim \left(\widehat{\pi}_{n-1}^{(i)} \times q_n\right)(\cdot)$.
  - With probability

$$\alpha_n(\mathbf{X}_n^{(i-1)}, \mathbf{X}_n^{*(i)}) = 1 \wedge \frac{w_n\left(\mathbf{X}_n^{*(i)}\right)}{w_n\left(\mathbf{X}_n^{(i-1)}\right)} \tag{2.4}$$

set $\mathbf{X}_n^{(i)} = \mathbf{X}_n^{*(i)}$, otherwise set $\mathbf{X}_n^{(i)} = \mathbf{X}_n^{(i-1)}$.

In this algorithm, $\widehat{\pi}_n^{(i)}$ is the empirical measure approximation of the target distribution π_n given by

$$\widehat{\pi}_n^{(i)}(d\mathbf{x}_n) = \frac{1}{i+1}\sum_{m=0}^{i} \delta_{\mathbf{X}_n^{(m)}}(d\mathbf{x}_n). \tag{2.5}$$

The (ratio of) normalizing constants can easily be estimated by

$$\widehat{Z_1}^{(i)} = \frac{1}{i}\sum_{m=1}^{i} w_1\left(\mathbf{X}_1^{*(m)}\right),$$

$$\left(\widehat{\frac{Z_n}{Z_{n-1}}}\right)^{(i)} = \frac{1}{i}\sum_{m=1}^{i} w_n\left(\mathbf{X}_n^{*(m)}\right). \tag{2.6}$$

Equation (2.6) follows from the identity

$$\frac{Z_n}{Z_{n-1}} = \int w_n(\mathbf{x}_n)(\pi_{n-1}\times q_n)(d\mathbf{x}_n)$$

and the fact that asymptotically (as $i \to \infty$) $\mathbf{X}_n^{*(i)}$ is distributed according to $(\pi_{n-1}\times q_n)(\mathbf{x}_n)$.

2.2. *Algorithm Settings.* Similarly to SMC methods, the performance of the SIMCMC algorithm depends heavily on the selection of the proposal distributions. However, it is possible to devise some useful guidelines for this sequence of (pseudo-)independent samplers, using reasoning similar to that adopted in the SMC framework. Asymptotically, $\mathbf{X}_n^{*(i)}$ is distributed according to $(\pi_{n-1}\times q_n)(\mathbf{x}_n)$ and $w_n(\mathbf{x}_n)$ is just the importance weight (up to a normalizing constant) between $\pi_n(\mathbf{x}_n)$ and $(\pi_{n-1}\times q_n)(\mathbf{x}_n)$. The proposal distribution minimizing the variance of this importance weight is simply given by

$$q_n^{\text{opt}}\left(\mathbf{x}_{n-1}, x_n\right) = \overline{\pi}_n\left(\mathbf{x}_{n-1}, x_n\right) \tag{2.7}$$

where $\overline{\pi}_n\left(\mathbf{x}_{n-1}, x_n\right)$ is the conditional density of x_n given $\mathbf{x}_{n-1}$ under π_n, that is

$$\overline{\pi}_n\left(\mathbf{x}_{n-1}, x_n\right) = \frac{\pi_n\left(\mathbf{x}_n\right)}{\pi_n\left(\mathbf{x}_{n-1}\right)}. \tag{2.8}$$

This yields

$$w_n^{\text{opt}}\left(\mathbf{x}_n\right) \propto \pi_{n/n-1}\left(\mathbf{x}_{n-1}\right) \tag{2.9}$$

where

$$\pi_{n/n-1}\left(\mathbf{x}_{n-1}\right) = \frac{\pi_n\left(\mathbf{x}_{n-1}\right)}{\pi_{n-1}\left(\mathbf{x}_{n-1}\right)} \tag{2.10}$$

with

$$\pi_n\left(\mathbf{x}_{n-1}\right) = \int_E \pi_n\left(\mathbf{x}_n\right) dx_n.$$

In this case, as $w_n^{\text{opt}}\left(\mathbf{x}_n\right)$ is independent of x_n, the algorithm described above can be further simplified. It is indeed possible to decide whether to accept or reject a candidate before even sampling it. This is more computationally efficient because if the move is to be rejected there is no need to sample the candidate. In most applications, it will be difficult to sample from (2.7) and/or to compute (2.9) as it involves computing $\pi_n\left(\mathbf{x}_{n-1}\right)$ up to a normalizing constant. In this case, we recommend approximating (2.7). Similar strategies have been developed successfully in the SMC framework [3], [9], [17], [20]. The advantages of such sampling strategies in the SIMCMC case will be demonstrated in the simulation section.

Generally speaking, most of the methodology developed in the SMC setting can be directly reapplied here. This includes the use of Rao-Blackwellisation techniques to reduce the dimensionality of the target distributions [9], [17] or of auxiliary particle-type ideas where we build target distributions biased towards 'promising' regions of the space [3], [20].

2.3. *Implementation Issues.*

2.3.1. *Burn-in and Storage requirements.* We have presented the algorithm without any burn-in. This can be easily included if necessary by considering at iteration i of the algorithm

$$\hat{\pi}_n^{(i)}(d\mathbf{x}_n) = \frac{1}{i+1-l(i,B)} \sum_{m=l(i,B)}^{i} \delta_{\mathbf{X}_n^{(m)}}(d\mathbf{x}_n),$$

where

$$l(i,B) = 0 \vee ((i-B) \wedge B)),$$

where B is an appropriate number of initial samples to be discarded as burn-in. Note that when $i \geq 2B$, we have $l(i,B) = B$.

Note that in its original form, the SIMCMC algorithm requires storing the sequences $\left\{\mathbf{X}_n^{(i)}\right\}_{n\in\mathbb{T}}$. This could be expensive if the number of target distributions P and/or the number of iterations of the SIMCMC are large. However, in many scenarios of interest including non-linear non-Gaussian state-space models or the scenarios considered in [7], it is possible to drastically reduce these storage requirements as we are only interested in estimating the marginals $\{\pi_n(x_n)\}$ and we have $w_n(\mathbf{x}_n) = w_n(x_{n-1}, x_n)$ and $q_n(\mathbf{x}_{n-1}, x_n) = q_n(x_{n-1}, x_n)$. In such cases, we only need to store $\left\{X_n^{(i)}\right\}_{n\in\mathbb{T}}$, resulting in significant memory savings.

2.3.2. *Combining Sampling Strategies.* In practice, we can combine the SIMCMC strategy with SMC methods; that is we can generate say N (approximate) samples from $\{\pi_n\}_{n\in\mathbb{T}}$ then we can use the SIMCMC strategy to increase the number of particles until the Monte Carlo estimates stabilize. We emphasize that SIMCMC will be primarily useful in the context where we do not have a predetermined computational budget. Indeed, if the computational budget is fixed, then we could also switch the iteration i and time n loops in the SIMCMC algorithm to obtain better estimates.

2.4. *Discussion and Extensions.* Standard MCMC methods do not address the problem solved by SIMCMC methods. Trans-dimensional MCMC methods [12] allow us to sample from a sequence of 'related' distributions but require the knowledge of the ratio of normalizing constants between different target distributions. and Simulated tempering and parallel tempering require all the target distributions to be defined on the same space and rely on MCMC kernels to explore each target distribution; see [15] for a recent discussion of such techniques. Ideas related to SIMCMC where a sequence of 'ideal' MCMC algorithms is approximated have recently appeared in physics [18] and statistics [14]. However, contrary to these algorithms, the target distributions considered here are of increasing dimension and the proposed interacting mechanism is simpler.

There are many possible extensions of the SIMCMC algorithm. In this respect the SIMCMC algorithm is somehow a proof-of-concept algorithm demonstrating that it is possible to make sequences targeting different distributions interact without the need to define a target distribution on an extended state-space. For example, instead of updating each chain sequentially, it is possible to update them in parallel using say standard MCMC updates and only attempting to 'jump' from π_{n-1} to π_n from time to time.

In the context of real-time applications where $\pi_n(\mathbf{x}_n)$ is typically the posterior distribution $p(\mathbf{x}_n | y_{1:n})$ of some states $\mathbf{x}_n$ given the observations $y_{1:n}$, SIMCMC methods can also be very useful. Indeed, SMC methods cannot easily address situations where the observations arrive at random times whereas SIMCMC methods allow us to make optimal use of what computing power is available by adding as many particles as possible until the arrival of a new observation. In such cases, a standard implementation would consist of updating our approximation of $\pi_n(\mathbf{x}_n)$ at 'time' n by adding iteratively particles to the approximations $\pi_{n-L+1}(\mathbf{x}_{n-L+1}), \ldots, \pi_{n-1}(\mathbf{x}_{n-1}), \pi_n(\mathbf{x}_n)$ for a lag $L \geq 1$ until the arrival of data y_{n+1}.

3. Convergence Results. We now present some convergence results for SIMCMC. Despite the non-Markovian nature of SIMCMC, we are here able to provide realistic verifiable assumptions ensuring the asymptotic consistency of the algorithm whereas a general result ensuring the convergence of the algorithm proposed in [18] and [14] has not yet been established.

Let us introduce $B(E_n) = \{f_n : E_n \to \mathbb{R} \text{ such that } \|f_n\| \leq 1\}$ where $\|f_n\| = \sup_{\mathbf{x}_n \in E_n} |f_n(\mathbf{x}_n)|$. We denote by $\mathbb{E}_{\mathbf{x}_{1:n}^{(0)}}[\cdot]$ the expectation with respect to the distribution of the simulated sequences initialized at $\mathbf{x}_{1:n}^{(0)} := \left(\mathbf{x}_1^{(0)}, \mathbf{x}_2^{(0)}, ..., \mathbf{x}_n^{(0)}\right)$ and $\mathbb{N}_0 = \mathbb{N} \cup \{0\}$. For any measure μ and test function f, we write $\mu(f) = \int \mu(dx) f(x)$.

We introduce $\mathcal{Q}_1 = \{\mathbf{x}_1 \in E_1 : q_1(\mathbf{x}_1) > 0\}$ and for $n \geq 2$ $\mathcal{Q}_n = \{\mathbf{x}_n \in E_n : (\pi_{n-1} \times q_n)(\mathbf{x}_n) > 0\}$. Our key assumption is the following.

Assumption A1. For any $n \in \mathbb{T}$, we have $\mathcal{S}_n \subseteq \mathcal{Q}_n$ and there exists $B_n < \infty$ such that for any $\mathbf{x}_n \in \mathcal{S}_n$

$$w_n(\mathbf{x}_n) \leq B_n. \tag{3.1}$$

This assumption is quite weak and can be easily checked in all the examples presented in Section 4. Note that a similar assumption also appears when $\mathbb{L}_p$ bounds are established for SMC methods [6].

Our first result establishes the convergence of the empirical averages towards the correct expectations at the standard Monte Carlo rate.

THEOREM 3.1. *Assume A1. For any $n \in \mathbb{T}$ and any $p \geq 1$ there exist $C_{1,n}, C_{2,p} < \infty$ such that for any $\mathbf{x}_{1:n}^{(0)} \in \mathcal{S}_1 \times \cdots \times \mathcal{S}_n$, $f_n \in B(E_n)$ and $i \in \mathbb{N}_0$*

$$\mathbb{E}_{\mathbf{x}_{1:n}^{(0)}}\left[\left|\hat{\pi}_n^{(i)}(f_n) - \pi_n(f_n)\right|^p\right]^{1/p} \leq \frac{C_{1,n}C_{2,p}}{(i+1)^{\frac{1}{2}}}.$$

By the Borel-Cantelli lemma, this also ensures almost sure convergence of the empirical averages. Our second result establishes that each sequence $\left\{\mathbf{X}_n^{(i)}\right\}$ converges towards π_n.

THEOREM 3.2. *Assume A1. For any $n \in \mathbb{T}$, $\mathbf{x}_{1:n}^{(0)} \in \mathcal{S}_1 \times \cdots \times \mathcal{S}_n$ and $f_n \in B(E_n)$ we have*

$$\lim_{i\to\infty} \mathbb{E}_{\mathbf{x}_{1:n}^{(0)}}\left[f_n\left(\mathbf{X}_n^{(i)}\right)\right] = \pi_n(f_n).$$

4. Applications. In this section, we will focus on the applications of SIMCMC to non-linear non-Gaussian state-space models. Consider an unobserved $E-$valued Markov process $\{X_n\}_{n\in\mathbb{T}}$ satisfying

$$X_1 \sim \mu(\cdot), \ \ X_n|\, X_{n-1} = x \sim f(x, \cdot).$$

We assume that we have access to observations $\{Y_n\}_{n\in\mathbb{T}}$ which, conditionally on $\{X_n\}$, are independent and distributed according to

$$Y_n|\,\{X_n = x\} \ \sim \ g(x, \cdot). \tag{4.1}$$

This family of models is important, because almost every stationary time series model appearing in the literature can be cast into this form. Given $y_{1:P}$, we are often interested in computing the sequence of posterior distributions $\{p(x_{1:n}|\, y_{1:n})\}_{n\in\mathbb{T}}$ to perform goodness-of-fit and/or to compute the marginal likelihood $p(y_{1:P})$. By defining the un-normalized distribution as

$$\gamma_n(x_{1:n}) = p(x_{1:n}, y_{1:n}) = \mu(x_1)\, g(x_1, y_1) \prod_{k=2}^{n} f(x_{k-1}, x_k)\, g(x_k, y_k) \tag{4.2}$$

(which is typically known pointwise), we have $\pi_n(\mathbf{x}_n) = p(x_{1:n}|\, y_{1:n})$ and $Z_n = p(y_{1:n})$ so that SIMCMC can be applied.

We will consider three examples here. In the first two cases, the SIMCMC algorithms are compared to their SMC counterparts. For a fixed number of iterations/particles, SMC and SIMCMC have approximately the same computational complexity. The same proposals and the same number of samples were thus used to allow for a fair comparison. Note that we chose not to use any burn-in period for the SIMCMC and we initialize the algorithm by picking $\mathbf{x}_n^{(0)} = \left(\mathbf{x}_{n-1}^{(0)}, x_n^{(0)}\right)$ for any n where $\mathbf{x}_P^{(0)}$ is a sample from the prior. The SMC algorithms were implemented using a stratified resampling procedure [13]. In the third example, we consider a slightly more complex problem of computing the likelihood of a stochastic volatility model with specified parameters. In this case the observation likelihood in fact depends on the entire past trajectory of the state, that is, we replace (4.1) by the assumption that $Y_n|\{X_{1:n} = x_{1:n}\} \sim g_n(x_{1:n}, \cdot)$.

4.1. *Linear Gaussian Model.* We consider a linear Gaussian model

$$\begin{aligned} X_n &= \phi X_{n-1} + V_n, \\ Y_n &= X_n + \sigma W_n \end{aligned} \tag{4.3}$$

with $X_1 \sim \mathcal{N}(0,1)$, $V_n \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0,1)$, $W_n \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0,1)$ and $\mathcal{N}(\mu, \sigma^2)$ is a Gaussian distribution of mean μ and variance σ^2. For this model we can compute the marginal likelihood $Z_P = p(y_{1:P})$ exactly using the Kalman filter. This allows us to compare our results to the ground truth.

We simulated a realization of $P = 100$ observations with $\phi = 0.95$ and $\sigma = 0.1$ for which $\log p(y_{1:P}) = -136.332$. We use two proposal distributions: the prior distribution $f(x_{n-1}, x_n)$ and the optimal distribution (4.3) given by $q_n(\mathbf{x}_{n-1}, x_n) \propto f(x_{n-1}, x_n)\, g(x_n, y_n)$. In both cases, it is easy to check that Assumption A1 is satisfied. In Figure 1, we display the estimates of $\log p(y_{1:P})$ obtained as a function of N for one realization of the SIMCMC algorithm.

In Table 1, we display the performance of both SIMCMC and SMC in terms of Root Mean Square Error (RMSE) for a varying number of samples and the two proposal distributions.

As expected, the RMSE of our estimates is drastically improved when the optimal distribution is used. For a small number of samples N, the performance of SMC is better than SIMCMC. This is not surprising as SIMCMC is an iterative MCMC-type algorithm and no burn in was used. For a larger number of samples, $N \geq 5000$, SMC and SIMCMC display very similar performance.

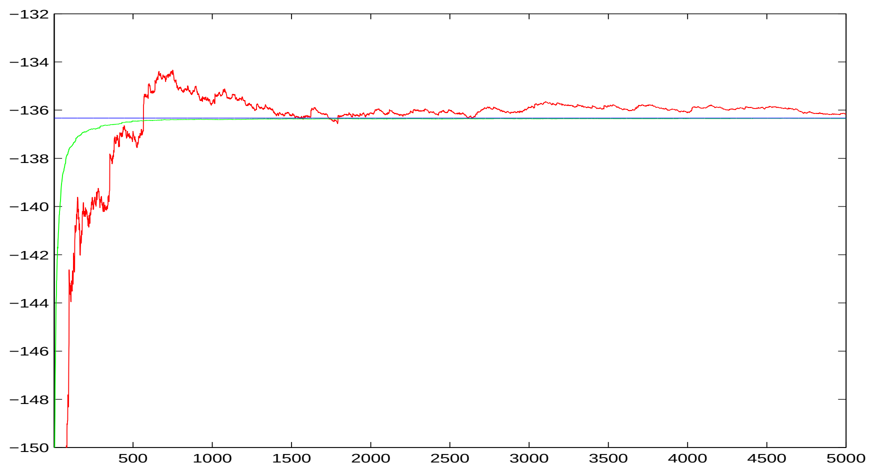

FIG 1. *True log-likelihood (blue), SIMCMC estimates computed using the prior proposal (red) and optimal proposal (green) as a function of N*

TABLE 1
RMSE of $\log p(y_{1:P})$ *over* 50 *realizations of the algorithms*

N	250	500	1000	2500	5000	10000	25000	50000
SIMCMC (prior)	8.20	3.09	2.39	1.10	0.64	0.46	0.23	0.17
SIMCMC (optimal)	0.32	0.11	0.09	0.05	0.03	0.02	0.01	0.01
SMC (prior)	11.80	4.13	1.97	0.94	0.68	0.40	0.29	0.19
SMC (optimal)	0.07	0.05	0.04	0.03	0.02	0.02	0.02	0.01

4.2. *A Nonlinear Non-Gaussian State-Space Model.* We now consider a nonlinear non-Gaussian state-space model introduced in [13] which has been used in many SMC publications

$$X_n = \frac{X_{n-1}}{2} + \frac{25X_{n-1}}{1+X_{n-1}^2} + 8\cos(1.2n) + V_n,$$
$$Y_n = \frac{X_n^2}{20} + W_n$$

where $X_1 \sim \mathcal{N}(0,5)$, $V_n \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0,25)$ and $W_n \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0,1)$. As the sign of the state X_n is not observed, the marginal posterior distributions $\{p(x_n|\, y_{1:n})\}$ are often bimodal. SMC approximations are able to capture properly the bimodality of the posteriors. This allows us to assess here whether SIMCMC can also explore properly these multimodal distributions by comparing SIMCMC to SMC results. The proposal distribution used is based on an extended Kalman filter and we checked that Assumption A1 was satisfied; see [9] for details.

TABLE 2
Standard Deviation of $\log p(y_{1:P})$ *over* 50 *realizations*

N	250	500	1000	2500	5000	10000	25000	50000
SIMCMC	4.71	3.60	1.74	1.37	0.87	0.78	0.54	0.38
SMC	1.61	1.52	1.14	0.94	0.63	0.58	0.47	0.30

We simulated a realization of $P = 100$ observations. In Table 2, we display the performance of both SIMCMC and SMC for a varying number of particles.

This model is more complex than the linear Gaussian model described earlier and the posterior distributions we are sampling can be multimodal. Consequently, the standard deviation of the estimates is not surprisingly higher than in the linear Gaussian case but the conclusions are similar. For a small number of samples, SMC outperforms SIMCMC whereas for a large number of samples the performance of both algorithms are comparable.

4.3. *Stochastic Volatility Model.* Stochastic volatility models are of considerable interest in the field of finance, having important applications in option pricing and (in the multivariate case) portfolio management. However, inference for these models remains a difficult computational problem because of the high-dimensional latent variable space that must be integrated out somehow. A number of authors have proposed MCMC and related methods, see for example [4] and references therein. It is also possible to use our SIMCMC approach for this purpose so as to compute the log-likelihood of the model for any particular specified parameter vector. This could in turn be used within a standard numerical optimization scheme or an MCMC scheme for parameter estimation, but in its own right also provides a convenient way to assess model goodness-of-fit.

We consider a particular form of a stochastic volatility model given in [1]. The log of a share price $\{\log S_t,\ t \in \mathbb{R}^+\}$ is assumed to satisfy the stochastic differential equation

$$d \log S(t) = [\mu + \beta\sigma^2(t)]dt + \sigma(t)dW(t) \tag{4.4}$$

$$d\sigma^2(t) = a\sigma^2(t)dt + bdL(t), \tag{4.5}$$

where μ, β, a and b are real-valued constants, and $\{W(t)\}$ and $\{L(t)\}$ are, respectively, standard Brownian motion and Poisson processes. $\{W(t)\}$ and $\{L(t)\}$ are independent of each other, and $\{L(t)\}$ is assumed to have rate λ. To ensure stationarity of the variance process $\{\sigma^2(t)\}$, we require $a < 0$. We also set $\sigma^2(0) = 0$.

This a continuous-time model for $S(t)$, but we will assume that observations are made at discrete times $n = 1, 2, \ldots, T$, and, as our data, we have the log-returns

$$Y_n = \log(S(n)/S(n-1)).$$

From (4.4) above, we see that

$$Y_n = \int_{n-1}^{n} \left[\mu + \beta\sigma^2(t)\right] dt + W_n \tag{4.6}$$

where

$$W_n = \int_{n-1}^{n} \sigma(t)\, dW(t).$$

It follows directly from (4.6) that conditionally on the segment of the volatility trajectory $\{\sigma^2(t),\ n-1 \le t < n\}$, Y_n is normally distributed, with mean $\mu + \beta\kappa_n$, and variance κ_n where

$$\kappa_n = \int_{n-1}^{n} \sigma^2(t)\, dt.$$

Furthermore, the segment $\{\sigma^2(t),\ n-1 \le t < n\}$ is completely determined by the (almost-surely finite) list of jump times of $\{L(t)\}$ in the interval $\{0 \le s < n\}$. Let us denote the set of jump times in the interval $[n-1, n)$ by

$$J_n = \{J_1^{(n)}, \ldots, J_{M_n}^{(n)}\}.$$

Then

$$\sigma^2(J_{k+1}^{(n)}) = \exp\left(a\left(J_{k+1}^{(n)} - J_k^{(n)}\right)\right)\sigma^2(J_k^{(n)}) + b, \text{ for } k = 0, ..., M_{n-1}$$

and

$$\begin{aligned}\kappa_n &= \sum_{k=0}^{M_n} \sigma^2(J_k^{(n)}) \int_{J_k^{(n)}}^{J_{k+1}^{(n)}} \exp[a(t - J_k^{(n)})]dt \\ &= \sum_{k=0}^{M_n} \sigma^2(J_k^{(n)}) \cdot \frac{\exp[a(J_{k+1}^{(n)} - J_k^{(n)})] - 1}{a}.\end{aligned} \tag{4.7}$$

Thus the observations under this model can be seen to satisfy the model

$$\begin{aligned} Y_n &\sim \mathcal{N}(\mu + \beta\kappa_n, \kappa_n) \\ X_n &= J_n, \end{aligned}$$

where J_n is the set of jump-times in $[n-1, n)$ as defined above. From (4.7), it is obvious that κ_n is a function of $X_1, \ldots, X_n$, and it is clear, because Poisson processes are "memoryless", that $\{X_n,\ n = 1, 2, \ldots\}$ is indeed a Markov process.

To illustrate the performance of the SIMCMC algorithm, we evaluate the log-likelihood of this model, fit to daily closing prices of National Australia Bank shares as listed on the Australian Stock Exchange, from Feb. 9th, 2002 to Aug. 10th, 2004. The prices and the corresponding log-returns are shown, respectively, in Figures 2 and 3, respectively.

FIG 2. *National Australia Bank share closing prices from Feb. 9th 2002 to Aug. 10th, 2004.*

For this data, we have $P = 506$ log-returns $y_1, \ldots, y_{506}$. For our proposal distributions $q_n(\mathbf{x}_{n-1}, x_n)$, we simply use the distribution of the arrivals in a Poisson process with rate 2λ. (Note that we choose to use proposals with a higher rate than the model because in general it is safer to use proposals with a higher spread than the target distribution than to use proposals with a smaller variance, but in this example, the factor of 2 makes little difference to the results obtained with a factor of 1.)

Figure 4 shows a realization of a sequence of estimates $\log p(y_{1:506})$ (i) for 25,000 iterations of the SIMCMC algorithm, applied using the burn-in adjustment of Subsection 2.3.1 with $B = 5000$. Results were obtained with parameters $a = -0.8$, $b = \exp(-7.0)$, $\lambda = \exp(-1.8)$, $\mu = 0$, $\delta = -0.5$. Note that these parameters are not optimal, they were simply obtained by simple hand-exploration of the parameter space over several iterations to obtain "reasonable" values. (For a more careful analysis, one could plug the SIMCMC algorithm into a numerical optimizer to obtain better estimates.) As a

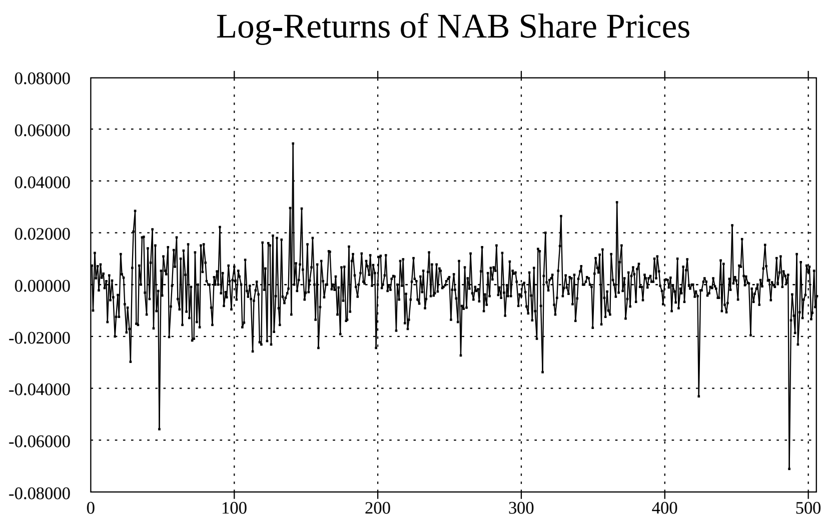

FIG 3. *Log-returns of the National Australia Bank share closing prices.*

base-line we also show (horizontal dashed line) the log-likelihood of the best iid-normal model for the log-returns. Not surprisingly, the stochastic volatility model, even with sub-optimal hand-chosen parameters, out-performs the iid-normal model significantly.

5. Discussion. We have described an iterative algorithm based on interacting non-Markovian sequences which is an alternative to SMC and have established convergence results validating this methodology. The algorithm is straightforward to implement for people already familiar with MCMC. For a fixed computational complexity, our simulation results indicate that, as expected, SMC usually outperforms SIMCMC for a small number of samples but that the performance of SMC and SIMCMC are very comparable when a large number of samples is used to obtain precise estimates of the quantities of interest. The main strength of SIMCMC is that it allows us to iteratively improve our estimates in an MCMC-like fashion until a suitable stopping criterion is met. This is useful as in numerous applications the number of particles required to ensure the estimates are reasonably precise is unknown. It is also useful in real-time applications when one is unsure of exactly how much time will be available between successive arrivals of data points.

As discussed in Subsection 2.4, numerous variants of SIMCMC can be

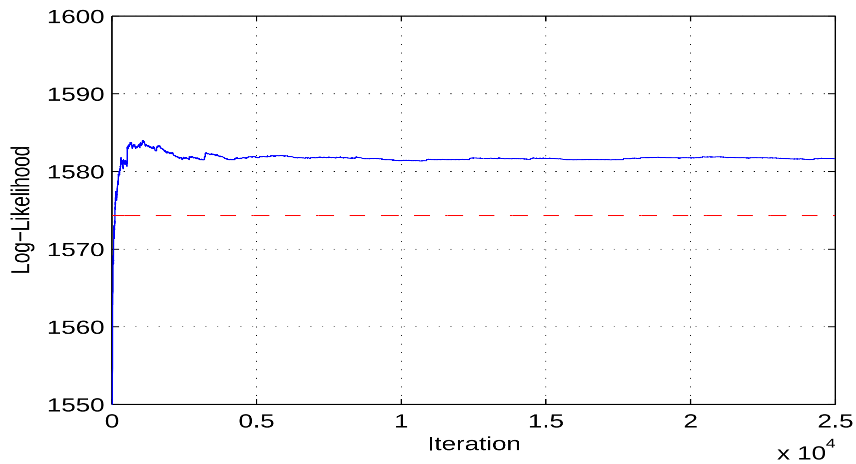

Fig 4. *Sequential estimates of the log-likelihood of the stochastic volatility model (4.4)-(4.5) for the National Australia Bank log-returns, over 25000 iterations of the SIMCMC algorithm. Parameters were not optimized, they were simply adjusted several times by hand to be "reasonable". Even so, the log-likelihood exceeds that of the maximum likelihood iid-normal model for log-returns, shown with a dashed line.*

easily developed which have no SMC equivalent. The fact that such schemes do not need to define a target distribution on an extended state-space admitting $\{\pi_n\}_{n\in\mathbb{T}}$ as marginals offers indeed a lot of flexibility. For example, if we have access to multiple processors, it is possible to sample from each π_n independently using standard MCMC and perform several interactions simultaneously; i.e. chains 1 and 2 could interact at the same time chains 3 and 4 interact. Adaptive versions of the algorithms can also be proposed by monitoring the acceptance ratio of the MH steps. If the acceptance probability of the MH move between say π_{n-1} and π_n is low, we could for example increase the number of proposals at this time index.

From a theoretical point of view, there are a number of interesting questions to explore. Under additional stability assumptions on the Feynman-Kac semigroup induced by $\{\pi_n\}_{n\in\mathbb{T}}$ and $\{q_n\}_{n\in\mathbb{T}}$ [6, chapter 4], we believe that it should be possible to obtain convergence results similar to [6, chapter 7] in an SMC framework ensuring that, for functions of the form $f_n(\mathbf{x}_n) = f_n(x_n)$, the bound $C_{1,n}$ in Theorem 3.1 is independent of n. We also conjecture that assumption A1 is not only sufficient but necessary.

APPENDIX A: PROOFS OF CONVERGENCE

The proof is inspired by ideas developed in [8]. However, in our context, it is possible to establish much stronger results than in this reference as we can characterize exactly the invariant distribution of some of the Markov kernels appearing in the analysis; see Proposition 2.

A.1. Notation. We denote by $\mathcal{P}(E_n)$ the set of probability measures on $(E_n, \mathcal{F}_n)$. We introduce the independent Metropolis-Hastings (MH) kernel $K_1 : E_1 \times \mathcal{F}_1 \rightarrow [0,1]$ defined by

$$K_1(\mathbf{x}_1, d\mathbf{x}_1') = \alpha_1(\mathbf{x}_1, \mathbf{x}_1') q_1(d\mathbf{x}_1') + \left(1 - \int \alpha_1(\mathbf{x}_1, \mathbf{y}_1) q_1(d\mathbf{y}_1)\right) \delta_{\mathbf{x}_1}(d\mathbf{x}_1'). \tag{A.1}$$

For $n \in \{2, ..., P\}$, we associate with any $\mu_{n-1} \in \mathcal{P}(E_{n-1})$ the Markov kernel $K_{n,\mu_{n-1}} : E_n \times \mathcal{F}_n \rightarrow [0,1]$ defined by

$$K_{n,\mu_{n-1}}(\mathbf{x}_n, d\mathbf{x}_n') = \alpha_n(\mathbf{x}_n, \mathbf{x}_n')(\mu_{n-1} \times q_n)(d\mathbf{x}_n') + \left(1 - \int \alpha_n(\mathbf{x}_n, \mathbf{y}_n)(\mu_{n-1} \times q_n)(d\mathbf{y}_n)\right) \delta_{\mathbf{x}_n}(d\mathbf{x}_n') \tag{A.2}$$

where $\mathbf{x}_n' = (\mathbf{x}_{n-1}', x_n')$. In (A.1) and (A.2), we have for $n \in \mathbb{T}$

$$\alpha_n(\mathbf{x}_n, \mathbf{x}_n') = 1 \wedge \frac{w_n(\mathbf{x}_n')}{w_n(\mathbf{x}_n)}.$$

We use $\|\cdot\|_{\text{tv}}$ to denote the total variation norm and for any Markov kernel

$$K^i(\mathbf{x}, d\mathbf{x}') := \int K^{i-1}(\mathbf{x}, d\mathbf{y}) K(\mathbf{y}, d\mathbf{x}').$$

A.2. Preliminary Results. We establish here the expression of the invariant distributions of the kernels $K_1(\mathbf{x}_1, d\mathbf{x}_1')$ and $K_{n,\mu_{n-1}}(\mathbf{x}_n, d\mathbf{x}_n')$ and establish that these kernels are geometrically ergodic. We also provide some perturbation bounds for $K_{n,\mu_{n-1}}(\mathbf{x}_n, d\mathbf{x}_n')$ and its invariant distribution.

LEMMA 1. *Assume A1. $K_1(\mathbf{x}_1, d\mathbf{x}_1')$ is uniformly geometrically ergodic of invariant distribution $\pi_1(d\mathbf{x}_1)$.*

By construction, $K_1(\mathbf{x}_1, d\mathbf{x}_1')$ is an MH algorithm of invariant distribution $\pi_1(d\mathbf{x}_1)$. Uniform ergodicity follows from A1; see for example Theorem 2.1. in [19].

Proposition 2. *Assume A1. For any* $n \in \{2, ..., P\}$ *and any* $\mu_{n-1} \in \mathcal{P}(E_{n-1})$, $K_{n,\mu_{n-1}}(\mathbf{x}_n, d\mathbf{x}'_n)$ *is uniformly geometrically ergodic of invariant distribution*

$$\omega_n(\mu_{n-1})(d\mathbf{x}_n) = \frac{\pi_{n/n-1}(\mathbf{x}_{n-1}) . (\mu_{n-1} \times \overline{\pi}_n)(d\mathbf{x}_n)}{\mu_{n-1}\left(\pi_{n/n-1}\right)} \tag{A.3}$$

where $\overline{\pi}_n(\mathbf{x}_{n-1}, dx_n)$ *and* $\pi_{n/n-1}(\mathbf{x}_{n-1})$ *are defined respectively in (2.8) and (2.10).*

Proof. To establish the result, it is sufficient to rewrite

$$\begin{aligned} w_n(\mathbf{x}_n) &= \frac{Z_n}{Z_{n-1}} \frac{\frac{\pi_n(\mathbf{x}_n)}{\pi_{n-1}(\mathbf{x}_{n-1})}\mu_{n-1}(\mathbf{x}_{n-1})}{(\mu_{n-1} \times q_n)(\mathbf{x}_n)} \\ &= \frac{Z_n}{Z_{n-1}} \frac{\pi_{n/n-1}(\mathbf{x}_{n-1})(\mu_{n-1} \times \overline{\pi}_n)(\mathbf{x}_n)}{(\mu_{n-1} \times q_n)(\mathbf{x}_n)}. \end{aligned}$$

This shows that $K_{n,\mu_{n-1}}(\mathbf{x}_n, d\mathbf{x}'_n)$ is nothing but a standard MH algorithm of proposal distribution $(\mu_{n-1} \times q_n)(\mathbf{x}_n)$ and target distribution given by (A.3). This distribution is always proper because A1 implies that $\pi_{n/n-1}(\mathbf{x}_{n-1}) < \infty$ over E_{n-1}. Uniform ergodicity follows from Theorem 2.1. in [19].■

Corollary. It follows that for any $n \in \{2, ..., P\}$ there exists $\rho_n < 1$ such that for any $m \in \mathbb{N}_0$ and $\mathbf{x}_n \in E_n$

$$\left\| K^m_{n,\mu_{n-1}}(\mathbf{x}_n, \cdot) - \omega_n(\mu_{n-1})(\cdot) \right\|_{\text{tv}} \leq \rho_n^m. \tag{A.4}$$

Proposition 3. *Assume A1. For any* $n \in \{2, ..., P\}$, *we have for any* μ_{n-1},$\nu_{n-1} \in \mathcal{P}(E_{n-1})$, $\mathbf{x}_n \in E_n$ *and* $m \in \mathbb{N}$

$$\left\| K^m_{n,\mu_{n-1}}(\mathbf{x}_n, \cdot) - K^m_{n,\nu_{n-1}}(\mathbf{x}_n, \cdot) \right\|_{\text{tv}} \leq \frac{2}{1-\rho_n} \left\| \mu_{n-1} - \nu_{n-1} \right\|_{\text{tv}} \tag{A.5}$$

and

$$\left\| \omega_n(\mu_{n-1}) - \omega_n(\nu_{n-1}) \right\|_{\text{tv}} \leq \frac{2}{1-\rho_n} \left\| \mu_{n-1} - \nu_{n-1} \right\|_{\text{tv}}. \tag{A.6}$$

Proof. For any $f_n \in B(E_n)$, we have the following decomposition

$$\begin{aligned} &K^m_{n,\mu_{n-1}}(f_n)(\mathbf{x}_n) - K^m_{n,\nu_{n-1}}(f_n)(\mathbf{x}_n) \\ &= \sum_{j=0}^{m-1} K^j_{n,\mu_{n-1}}\left(K_{n,\mu_{n-1}} - K_{n,\nu_{n-1}}\right) K^{m-j-1}_{n,\nu_{n-1}}(f_n)(\mathbf{x}_n) \\ &= \sum_{j=0}^{m-1} K^j_{n,\mu_{n-1}}\left(K_{n,\mu_{n-1}} - K_{n,\nu_{n-1}}\right)\left(K^{m-j-1}_{n,\nu_{n-1}}(f_n)(\mathbf{x}_n) - \omega_n(\nu_{n-1})(f_n)\right). \end{aligned}$$

From A1, we know that for any $\nu_{n-1} \in \mathcal{P}(E_{n-1})$

$$\left\| K_{n,\nu_{n-1}}^{m-j-1}(\mathbf{x}_n, \cdot) - \omega_n(\nu_{n-1}) \right\|_{\text{tv}} \leq \rho_n^{m-j-1}$$

and from (A.2) for any $\mathbf{x}_n \in E_n$ and $f_n \in B(E_n)$

$$\begin{aligned}
& K_{n,\mu_{n-1}}(f_n)(\mathbf{x}_n) - K_{n,\nu_{n-1}}(f_n)(\mathbf{x}_n) \\
&= \int f_n(\mathbf{x}'_n)\, \alpha_n(\mathbf{x}_n, \mathbf{x}'_n)\, ((\mu_{n-1} - \nu_{n-1}) \times q_n)(d\mathbf{x}'_n) \\
&+ f_n(\mathbf{x}_n) \int \alpha_n(\mathbf{x}_n, \mathbf{y}'_n)\, ((\nu_{n-1} - \mu_{n-1}) \times q_n)(d\mathbf{y}'_n)
\end{aligned}$$

thus

$$\left\| K_{n,\mu_{n-1}}(\mathbf{x}_n, \cdot) - K_{n,\nu_{n-1}}(\mathbf{x}_n, \cdot) \right\|_{\text{tv}} \leq 2 \left\| \mu_{n-1} - \nu_{n-1} \right\|_{\text{tv}}.$$

So

$$\begin{aligned}
\left\| K_{n,\mu_{n-1}}(\mathbf{x}_n, \cdot) - K_{n,\nu_{n-1}}(\mathbf{x}_n, \cdot) \right\|_{\text{tv}} &\leq 2 \left\| \mu_{n-1} - \nu_{n-1} \right\|_{\text{tv}} \sum_{j=0}^{m-1} \rho_n^{m-j-1} \\
&= 2 \frac{1 - \rho_n^m}{1 - \rho_n} \left\| \mu_{n-1} - \nu_{n-1} \right\|_{\text{tv}}.
\end{aligned}$$

Hence (A.5) follows and we obtain (A.6) by taking the limit as $m \to \infty$.■

A.3. Convergence of Averages. For any $n \in \{2, ..., P\}$, $p \geq 1$ and $f_n \in B(E_n)$ we want to study

$$\mathbb{E}_{\mathbf{x}_{1:n}^{(0)}} \left[\left| \widehat{\pi}_n^{(i)}(f_n) - \pi_n(f_n) \right|^p \right]^{1/p}.$$

We have

$$\widehat{\pi}_n^{(i)}(f_n) - \pi_n(f_n) = \widehat{\pi}_n^{(i)}(f_n) - S_n^{(i)}(f_n) + S_n^{(i)}(f_n) - \pi_n(f_n) \tag{A.7}$$

where

$$S_n^{(i)}(f_n) = \frac{1}{i+1} \sum_{m=0}^{i} \omega_n \left(\widehat{\pi}_{n-1}^{(m)} \right)(f_n).$$

To study the first term on the rhs of (A.7), we introduce the Poisson equation

$$f_n(x) - \omega_n(\mu)(f_n) = \widehat{f}_{n,\mu}(x) - K_{n,\mu}\left(\widehat{f}_{n,\mu} \right)(x)$$

whose solution, if $K_{n,\mu}$ is geometrically ergodic, is given by

$$\widehat{f}_{n,\mu}(x) = \sum_{i \in \mathbb{N}_0} \left[K_{n,\mu}^i(f_n)(x) - \omega_n(\mu)(f_n) \right]. \tag{A.8}$$

We have

(A.9)

$$(i+1)\left[\widehat{\pi}_n^{(i)}(f_n) - S_n^{(i)}\right] = M_n^{(i+1)}(f_n) + \sum_{m=0}^{i}\left[\widehat{f}_{n,\widehat{\pi}_{n-1}^{(m+1)}}\left(\mathbf{X}_n^{(m+1)}\right) - \widehat{f}_{n,\widehat{\pi}_{n-1}^{(m)}}\left(\mathbf{X}_n^{(m+1)}\right)\right] + \widehat{f}_{n,\widehat{\pi}_{n-1}^{(0)}}\left(\mathbf{X}_n^{(0)}\right) - \widehat{f}_{n,\widehat{\pi}_{n-1}^{(i+1)}}\left(\mathbf{X}_n^{(i+1)}\right)$$

where

$$M_n^{(i)}(f_n) = \sum_{m=0}^{i-1}\left[\widehat{f}_{n,\widehat{\pi}_{n-1}^{(m)}}\left(\mathbf{X}_n^{(m+1)}\right) - K_{\widehat{\pi}_{n-1}^{(m)}}\left(\widehat{f}_{n,\widehat{\pi}_{n-1}^{(m)}}\right)\left(\mathbf{X}_n^{(m)}\right)\right] \tag{A.10}$$

is a $\mathcal{G}_n^i$-martingale with $\mathcal{G}_n^i = \sigma\left(\mathbf{X}_1^{(1:i)}, \mathbf{X}_2^{(1:i)}, ..., \mathbf{X}_n^{(1:i)}\right)$ where we define $\mathbf{X}_k^{(1:i)} = \left(\mathbf{X}_k^{(1)}, ..., \mathbf{X}_k^{(i)}\right)$.

We remind the reader that $B(E_n) = \{f_n : E_n \to \mathbb{R} \text{ such that } \|f_n\| \leq 1\}$ where $\|f_n\| = \sup_{\mathbf{x}_n \in E_n} |f_n(\mathbf{x}_n)|$. We establish the following propositions.

PROPOSITION 4. *Assume A1. For any $n \in \{2, ..., P\}$, $\mathbf{x}_{1:n}^{(0)}$, $p \geq 1$, $f_n \in B(E_n)$ and $m \in \mathbb{N}_0$, we have*

$$\mathbb{E}_{\mathbf{x}_{1:n}^{(0)}}\left[\left|\widehat{f}_{n,\widehat{\pi}_{n-1}^{(m)}}\left(\mathbf{X}_n^{(m+1)}\right)\right|^p\right]^{1/p} \leq \frac{1}{1-\rho_n}.$$

Proof. Assumption A1 ensures that $\widehat{f}_{n,\widehat{\pi}_{n-1}^{(m)}}$ is given by an expression of the form (A.8). We have

$$\begin{aligned}
&\mathbb{E}_{\mathbf{x}_{1:n}^{(0)}}\left[\left|\widehat{f}_{n,\widehat{\pi}_{n-1}^{(m)}}\left(\mathbf{X}_n^{(m+1)}\right)\right|^p\right]^{1/p} \\
&\leq \sum_{i\in\mathbb{N}_0} \mathbb{E}_{\mathbf{x}_{1:n}^{(0)}}\left[\left|K^i_{n,\widehat{\pi}_{n-1}^{(m)}}(f_n)\left(\mathbf{X}_n^{(m+1)}\right) - \omega_n\left(\widehat{\pi}_{n-1}^{(m)}\right)(f_n)\right|^p\right]^{1/p} \\
&\leq \sum_{i\in\mathbb{N}_0} \mathbb{E}_{\mathbf{x}_{1:n}^{(0)}}\left[\mathbb{E}_{\mathbf{x}_{1:n}^{(0)}}\left(\left|K^i_{n,\widehat{\pi}_{n-1}^{(m)}}(f_n)\left(\mathbf{X}_n^{(m+1)}\right) - \omega_n\left(\widehat{\pi}_{n-1}^{(m)}\right)(f_n)\right|^p \middle| \mathcal{G}_{n-1}^m\right)\right]^{1/p} \\
&\leq \sum_{i\in\mathbb{N}_0} \rho_n^i = \frac{1}{1-\rho_n}.
\end{aligned}$$

using Minkowski's inequality and the fact that $K_{n,\widehat{\pi}_{n-1}^{(m)}}$ is an uniformly ergodic Markov kernel conditional upon $\mathcal{G}_{n-1}^m$ using A1.■

PROPOSITION 5. *Assume A1. For any $n \in \{2, ..., P\}$ and any $p \geq 1$, there exist $B_{1,n}, B_{2,p} < \infty$ such that for any $\mathbf{x}_{1:n}^{(0)}$, $f_n \in B(E_n)$ and $m \in \mathbb{N}$*

$$\mathbb{E}_{\mathbf{x}_{1:n}^{(0)}}\left[\left|M_n^{(m)}(f_n)\right|^p\right]^{1/p} \leq B_{1,n}B_{2,p}\, m^{\frac{1}{2}}.$$

Proof. For $p > 1$ we use Burkholder's inequality (e.g. [21, p. 499]); i.e. there exist constants $C_{1,n}, C_{2,p} < \infty$ such that

$$\begin{aligned}(A.11)\quad & \mathbb{E}_{\mathbf{x}_{1:n}^{(0)}}\left[\left|M_n^{(m)}(f_n)\right|^p\right]^{1/p} \\ \leq\ & C_{1,n}C_{2,p}\mathbb{E}_{\mathbf{x}_{1:n}^{(0)}}\left[\left(\sum_{i=0}^{m-1}\left[\widehat{f}_{n,\widehat{\pi}_{n-1}^{(i)}}\left(\mathbf{X}_n^{(i+1)}\right) - K_{n,\widehat{\pi}_{n-1}^{(i)}}\left(\widehat{f}_{n,\widehat{\pi}_{n-1}^{(i)}}\right)\left(\mathbf{X}_n^{(i)}\right)\right]^2\right)^{p/2}\right]^{1/p}.\end{aligned}$$

For $p \in (1, 2)$, we can bound the lhs of (A.11)

$$\begin{aligned}& \mathbb{E}_{\mathbf{x}_{1:n}^{(0)}}\left[\left(\sum_{i=0}^{m-1}\left[\widehat{f}_{n,\widehat{\pi}_{n-1}^{(i)}}\left(\mathbf{X}_n^{(i+1)}\right) - K_{n,\widehat{\pi}_{n-1}^{(i)}}\left(\widehat{f}_{n,\widehat{\pi}_{n-1}^{(i)}}\right)\left(\mathbf{X}_n^{(i)}\right)\right]^2\right)^{p/2}\right]^{1/p} \\ & \leq \mathbb{E}_{\mathbf{x}_{1:n}^{(0)}}\left[\left(2\sum_{i=0}^{m-1}\left[\left|\widehat{f}_{n,\widehat{\pi}_{n-1}^{(i)}}\left(\mathbf{X}_n^{(i+1)}\right)\right|^2 + \left|K_{n,\widehat{\pi}_{n-1}^{(i)}}\left(\widehat{f}_{n,\widehat{\pi}_{n-1}^{(i)}}\right)\left(\mathbf{X}_n^{(i)}\right)\right|^2\right]\right)^{p/2}\right]^{1/p} \\ & \leq \mathbb{E}_{\mathbf{x}_{1:n}^{(0)}}\left[\left(2\sum_{i=0}^{m-1}\left[\left|\widehat{f}_{n,\widehat{\pi}_{n-1}^{(i)}}\left(\mathbf{X}_n^{(i+1)}\right)\right|^2 + \left|K_{n,\widehat{\pi}_{n-1}^{(i)}}\left(\widehat{f}_{n,\widehat{\pi}_{n-1}^{(i)}}\right)\left(\mathbf{X}_n^{(i)}\right)\right|^2\right]\right)\right]^{1/2}\end{aligned}$$

using $(a-b)^2 \leq 2(a^2+b^2)$ and Jensen's inequality. Now using Jensen's inequality again, we have

$$\mathbb{E}_{\mathbf{x}_{1:n}^{(0)}}\left[\left|K_{n,\widehat{\pi}_{n-1}^{(i)}}\left(\widehat{f}_{n,\widehat{\pi}_{n-1}^{(i)}}\right)\left(\mathbf{X}_n^{(i)}\right)\right|^2\right] \leq \mathbb{E}_{\mathbf{x}_{1:n}^{(0)}}\left[K_{n,\widehat{\pi}_{n-1}^{(i)}}\left(\left|\widehat{f}_{n,\widehat{\pi}_{n-1}^{(i)}}\right|^2\right)\left(\mathbf{X}_n^{(i)}\right)\right]$$

and using Proposition 4, we obtain the bound

$$\mathbb{E}_{\mathbf{x}_{1:n}^{(0)}}\left[\left(\sum_{i=0}^{m-1}\left[\widehat{f}_{n,\widehat{\pi}_{n-1}^{(i)}}\left(\mathbf{X}_n^{(i+1)}\right) - K_{n,\widehat{\pi}_{n-1}^{(i)}}\left(\widehat{f}_{n,\widehat{\pi}_{n-1}^{(i)}}\right)\left(\mathbf{X}_n^{(i)}\right)\right]^2\right)^{p/2}\right]^{1/p} \leq \frac{2}{1-\rho_n}m^{\frac{1}{2}}.$$

For $p \geq 2$, we we can bound the lhs of (A.11) through Minkowski's inequality

$$\begin{aligned}& \mathbb{E}_{\mathbf{x}_{1:n}^{(0)}}\left[\left|M_n^{(m)}(f_n)\right|^p\right]^{1/p} \\ \leq\ & C_{1,n}C_{2,p}\left(\sum_{i=0}^{m-1}\mathbb{E}_{\mathbf{x}_{1:n}^{(0)}}\left[\left|\widehat{f}_{n,\widehat{\pi}_{n-1}^{(i)}}\left(\mathbf{X}_n^{(i+1)}\right) - K_{n,\widehat{\pi}_{n-1}^{(i)}}\left(\widehat{f}_{n,\widehat{\pi}_{n-1}^{(i)}}\right)\left(\mathbf{X}_n^{(i)}\right)\right|^p\right]^{2/p}\right)^{1/2}.\end{aligned}$$

Using Minkowski's inequality again

$$\mathbb{E}_{\mathbf{x}_{1:n}^{(0)}}\left[\left|\widehat{f}_{n,\widehat{\pi}_{n-1}^{(i)}}\left(\mathbf{X}_n^{(i+1)}\right)-K_{n,\widehat{\pi}_{n-1}^{(i)}}\left(\widehat{f}_{n,\widehat{\pi}_{n-1}^{(i)}}\right)\left(\mathbf{X}_n^{(i)}\right)\right|^p\right]$$
$$\leq\left(\mathbb{E}_{\mathbf{x}_{1:n}^{(0)}}\left[\left|\widehat{f}_{n,\widehat{\pi}_{n-1}^{(i)}}\left(\mathbf{X}_n^{(i+1)}\right)\right|^p\right]^{1/p}+\mathbb{E}_{\mathbf{x}_{1:n}^{(0)}}\left[\left|K_{n,\widehat{\pi}_{n-1}^{(i)}}\left(\widehat{f}_{n,\widehat{\pi}_{n-1}^{(i)}}\right)\left(\mathbf{X}_n^{(i)}\right)\right|^p\right]^{1/p}\right)^p.$$

Now from Proposition 4 and Jensen's inequality, we can conclude for $p \geq 1$. For $p = 1$, we use Davis' inequality (e.g. [21, p. 499]) to obtain the result using similar arguments which are not repeated here.■

PROPOSITION 6. *Assume A1. For any $n \in \{2, ..., P\}$ and $p \geq 1$ there exists $B_n < \infty$ such that for any $\mathbf{x}_{1:n}^{(0)}$, $f_n \in B(E_n)$ and $m \in \mathbb{N}_0$*

$$\mathbb{E}_{\mathbf{x}_{1:n}^{(0)}}\left[\left|\widehat{f}_{n,\widehat{\pi}_{n-1}^{(m+1)}}\left(\mathbf{X}_n^{(m+1)}\right)-\widehat{f}_{n,\widehat{\pi}_{n-1}^{(m)}}\left(\mathbf{X}_n^{(m+1)}\right)\right|^p\right]^{1/p}\leq\frac{B_n}{m+2}$$

Proof. Our proof is based on the following key decomposition established in Lemma 3.2. in [8]

(A.12)
$$\widehat{f}_{n,\widehat{\pi}_{n-1}^{(m+1)}}\left(\mathbf{X}_n^{(m+1)}\right)-\widehat{f}_{n,\widehat{\pi}_{n-1}^{(m)}}\left(\mathbf{X}_n^{(m+1)}\right)+\omega_n\left(\widehat{\pi}_{n-1}^{(m+1)}\right)\left(\widehat{f}_{n,\widehat{\pi}_{n-1}^{(m)}}\right)$$
$$=\sum_{i,j\in\mathbb{N}_0}\left(\delta_{\mathbf{X}_n^{(m+1)}}-\omega_n\left(\widehat{\pi}_{n-1}^{(m+1)}\right)\right)K^i_{n,\widehat{\pi}_{n-1}^{(m+1)}}\left(K_{n,\widehat{\pi}_{n-1}^{(m+1)}}-K_{n,\widehat{\pi}_{n-1}^{(m)}}\right)$$
$$\times K^j_{n,\widehat{\pi}_{n-1}^{(m)}}\left(f_n-\omega_n\left(\widehat{\pi}_{n-1}^{(m)}\right)(f_n)\right).$$

We have

$$\left|\left(\delta_{\mathbf{X}_n^{(m+1)}}-\omega_n\left(\widehat{\pi}_{n-1}^{(m+1)}\right)\right)K^i_{n,\widehat{\pi}_{n-1}^{(m+1)}}\left(K_{n,\widehat{\pi}_{n-1}^{(m+1)}}-K_{n,\widehat{\pi}_{n-1}^{(m)}}\right)K^j_{n,\widehat{\pi}_{n-1}^{(m)}}\left(f_n-\omega_n\left(\widehat{\pi}_{n-1}^{(m)}\right)(f_n)\right)\right|$$
$$=\left|\left(\delta_{\mathbf{X}_n^{(m+1)}}-\omega_n\left(\widehat{\pi}_{n-1}^{(m+1)}\right)\right)K^i_{n,\widehat{\pi}_{n-1}^{(m+1)}}\left(K_{n,\widehat{\pi}_{n-1}^{(m+1)}}-K_{n,\widehat{\pi}_{n-1}^{(m)}}\right)K^j_{n,\widehat{\pi}_{n-1}^{(m)}}(f_n)\right|$$
$$\leq\rho_n^j\left\|\left(\delta_{\mathbf{X}_n^{(m+1)}}-\omega_n\left(\widehat{\pi}_{n-1}^{(m+1)}\right)\right)K^i_{n,\widehat{\pi}_{n-1}^{(m+1)}}\left(K_{n,\widehat{\pi}_{n-1}^{(m+1)}}-K_{n,\widehat{\pi}_{n-1}^{(m)}}\right)\right\|_{\text{tv}}$$
$$\leq\rho_n^j\times\frac{2}{1-\rho_n}\left\|\widehat{\pi}_{n-1}^{(m+1)}-\widehat{\pi}_{n-1}^{(m)}\right\|_{\text{tv}}\left\|\left(\delta_{\mathbf{X}_n^{(m+1)}}-\omega_n\left(\widehat{\pi}_{n-1}^{(m+1)}\right)\right)K^i_{n,\widehat{\pi}_{n-1}^{(m+1)}}\right\|_{\text{tv}}$$
$$\leq\rho_n^j\times\frac{2}{1-\rho_n}\left\|\widehat{\pi}_{n-1}^{(m+1)}-\widehat{\pi}_{n-1}^{(m)}\right\|_{\text{tv}}\times\rho_n^i\left\|\delta_{\mathbf{X}_n^{(m+1)}}-\omega_n\left(\widehat{\pi}_{n-1}^{(m+1)}\right)\right\|_{\text{tv}}$$

(A.13)
$$\leq\frac{2\rho_n^{i+j}}{1-\rho_n}\left\|\widehat{\pi}_{n-1}^{(m+1)}-\widehat{\pi}_{n-1}^{(m)}\right\|_{\text{tv}}.$$

using A1, (A.5) in Proposition 3 and A1 again.

Now we have

$$\left|\omega_n\left(\widehat{\pi}_{n-1}^{(m+1)}\right)\left(\widehat{f}_{n,\widehat{\pi}_{n-1}^{(m)}}\right)\right| = \left|\omega_n\left(\widehat{\pi}_{n-1}^{(m+1)}\right)\left(\sum_{i\in\mathbb{N}_0}\left[K^i_{n,\widehat{\pi}_{n-1}^{(m)}}(f_n) - \omega_n\left(\widehat{\pi}_{n-1}^{(m)}\right)(f_n)\right]\right)\right|$$

$$= \sum_{i\in\mathbb{N}_0}\left|\left(\omega_n\left(\widehat{\pi}_{n-1}^{(m+1)}\right) - \omega_n\left(\widehat{\pi}_{n-1}^{(m)}\right)\right)K^i_{n,\widehat{\pi}_{n-1}^{(m)}}(f_n)\right|$$

$$\leq \sum_{i\in\mathbb{N}_0}\rho_n^i\left\|\omega_n\left(\widehat{\pi}_{n-1}^{(m+1)}\right) - \omega_n\left(\widehat{\pi}_{n-1}^{(m)}\right)\right\|_{\mathrm{tv}}$$

$$\leq \frac{2}{(1-\rho_n)^2}\left\|\widehat{\pi}_{n-1}^{(m+1)} - \widehat{\pi}_{n-1}^{(m)}\right\|_{\mathrm{tv}}. \tag{A.14}$$

using A1 and (A.6) in Proposition 3.

Now for any $f_{n-1} \in B(E_{n-1})$, we have

$$\widehat{\pi}_{n-1}^{(m+1)}(f_{n-1}) - \widehat{\pi}_{n-1}^{(m)}(f_{n-1}) = \frac{f_{n-1}\left(\mathbf{X}_{n-1}^{(m+1)}\right)}{m+2} - \frac{\widehat{\pi}_{n-1}^{(m)}(f_{n-1})}{m+2}.$$

thus

$$\left\|\widehat{\pi}_{n-1}^{(m+1)} - \widehat{\pi}_{n-1}^{(m)}\right\|_{\mathrm{tv}} \leq \frac{2}{m+2}. \tag{A.15}$$

The result follows now directly combining (A.12), (A.13), (A.14), (A.15) and using Minkowski's inequality.■

PROPOSITION 7. *Assume A1. For any $n \in \{2, ..., P\}$ and any $p \geq 1$ there exists $B_{1,n}, B_{2,p} < \infty$ such that for $\mathbf{x}_{1:n}^{(0)}$, $f_n \in B(E_n)$ and $i \in \mathbb{N}_0$*

$$\mathbb{E}_{\mathbf{x}_{1:n}^{(0)}}\left[\left|\widehat{\pi}_n^{(i)}(f_n) - S_n^{(i)}(f_n)\right|^p\right]^{1/p} \leq \frac{B_{1,n}B_{2,p}}{(i+1)^{\frac{1}{2}}}$$

Proof. Using (A.9) and Minkowski's inequality, we obtain

(A.16)

$$\mathbb{E}_{\mathbf{x}_{1:n}^{(0)}}\left[\left|\widehat{\pi}_n^{(i)}(f_n) - S_n^{(i)}(f_n)\right|^p\right]^{1/p}$$
$$\leq \frac{1}{(i+1)}\mathbb{E}_{\mathbf{x}_{1:n}^{(0)}}\left[\left|M_n^{(i+1)}(f_n)\right|^p\right]^{1/p}$$
$$+ \frac{1}{(i+1)}\sum_{m=0}^{i}\mathbb{E}_{\mathbf{x}_{1:n}^{(0)}}\left[\left|\widehat{f}_{n,\widehat{\pi}_{n-1}^{(m+1)}}\left(\mathbf{X}_n^{(m+1)}\right) - \widehat{f}_{n,\widehat{\pi}_{n-1}^{(m)}}\left(\mathbf{X}_n^{(m+1)}\right)\right|^p\right]^{1/p}$$
$$+ \frac{1}{i+1}\mathbb{E}_{\mathbf{x}_{1:n}^{(0)}}\left[\left|\widehat{f}_{n,\widehat{\pi}_{n-1}^{(0)}}\left(\mathbf{X}_n^{(0)}\right)\right|^p\right]^{1/p} + \frac{1}{i+1}\mathbb{E}_{\mathbf{x}_{1:n}^{(0)}}\left[\left|\widehat{f}_{n,\widehat{\pi}_{n-1}^{(i+1)}}\left(\mathbf{X}_n^{(i+1)}\right)\right|^p\right]^{1/p}.$$

The first term on the rhs of (A.16) is bounded using Proposition 5, the term on the last line of the rhs are going to zero because of Proposition 4. For the second term, we obtain using Proposition 6

$$\sum_{m=0}^{i} \mathbb{E}_{\mathbf{x}_{1:n}^{(0)}}\left[\left|\widehat{f}_{n,\widehat{\pi}_{n-1}^{(m+1)}}\left(\mathbf{X}_n^{(m+1)}\right)-\widehat{f}_{n,\widehat{\pi}_{n-1}^{(m)}}\left(\mathbf{X}_n^{(m+1)}\right)\right|^p\right]^{1/p} \leq \sum_{m=0}^{i}\frac{B_n}{m+2} \leq B_n \log(i+2)$$

The result follows.■

Proof of Theorem 3.1. Under A1, the result is clearly true for $n=1$ thanks to Lemma 1. Assume it is true for $n-1$ and let us prove it for n. We have using Minkowski's inequality

$$\mathbb{E}_{\mathbf{x}_{1:n}^{(0)}}\left[\left|\widehat{\pi}_n^{(i)}(f_n)-\pi_n(f_n)\right|^p\right]^{1/p} \leq \mathbb{E}_{\mathbf{x}_{1:n}^{(0)}}\left[\left|\widehat{\pi}_n^{(i)}(f_n)-S_n^{(i)}(f_n)\right|^p\right]^{1/p} + \mathbb{E}_{\mathbf{x}_{1:n}^{(0)}}\left[\left|S_n^{(i)}(f_n)-\pi_n(f_n)\right|^p\right]^{1/p}.$$

The first term on the rhs can be bounded using Proposition 7. For the second term, we have

$$\mathbb{E}_{\mathbf{x}_{1:n}^{(0)}}\left[\left|S_n^{(i)}(f_n)-\pi_n(f_n)\right|^p\right]^{1/p} \leq \frac{1}{(i+1)}\sum_{m=0}^{i}\mathbb{E}_{\mathbf{x}_{1:n}^{(0)}}\left[\left|\omega_n\left(\widehat{\pi}_{n-1}^{(m)}\right)(f_n)-\omega_n(\pi_{n-1})(f_n)\right|^p\right]^{1/p}.$$

Using (A.3), we obtain

$$\begin{aligned}
&\omega_n(\pi_{n-1})(f_n)-\omega_n\left(\widehat{\pi}_{n-1}^{(m)}\right)(f_n) \\
&= \frac{(\pi_{n-1}\times\overline{\pi}_n)\left(\pi_{n/n-1}.f_n\right)}{\pi_{n-1}\left(\pi_{n/n-1}\right)} - \frac{\left(\widehat{\pi}_{n-1}^{(m)}\times\overline{\pi}_n\right)\left(\pi_{n/n-1}.f_n\right)}{\widehat{\pi}_{n-1}^{(m)}\left(\pi_{n/n-1}\right)} \\
&= \frac{\left(\left(\pi_{n-1}-\widehat{\pi}_{n-1}^{(m)}\right)\times\overline{\pi}_n\right)\left(\pi_{n/n-1}.f_n\right).\widehat{\pi}_{n-1}^{(m)}\left(\pi_{n/n-1}\right)}{\widehat{\pi}_{n-1}^{(m)}\left(\pi_{n/n-1}\right).\pi_{n-1}\left(\pi_{n/n-1}\right)} \\
&+ \frac{\left(\widehat{\pi}_{n-1}^{(m)}\times\overline{\pi}_n\right)\left(\pi_{n/n-1}.f_n\right).\left(\widehat{\pi}_{n-1}^{(m)}-\pi_{n-1}\right)\left(\pi_{n/n-1}\right)}{\widehat{\pi}_{n-1}^{(m)}\left(\pi_{n/n-1}\right).\pi_{n-1}\left(\pi_{n/n-1}\right)}
\end{aligned}$$

so, as $\pi_{n-1}\left(\pi_{n/n-1}\right)=1$, we have

$$\begin{aligned}
&\left|\omega_n\left(\pi_{n-1}\right)\left(f_n\right)-\omega_n\left(\widehat{\pi}_{n-1}^{(m)}\right)\left(f_n\right)\right| \\
&\leq\left|\left(\left(\pi_{n-1}-\widehat{\pi}_{n-1}^{(m)}\right)\times\overline{\pi}_n\right)\left(\pi_{n/n-1}.f_n\right)\right| \\
&+\frac{\left|\left(\widehat{\pi}_{n-1}^{(m)}\times\overline{\pi}_n\right)\left(\pi_{n/n-1}.f_n\right).\left(\widehat{\pi}_{n-1}^{(m)}-\pi_{n-1}\right)\left(\pi_{n/n-1}\right)\right|}{\widehat{\pi}_{n-1}^{(m)}\left(\pi_{n/n-1}\right)}.
\end{aligned}$$

Assumption A1 implies that there exists $D_n<\infty$ such that $\pi_{n/n-1}\left(\mathbf{x}_{n-1}\right)<D_n$ over E_{n-1}. Thus we have using the induction hypothesis

$$\begin{aligned}
&\mathbb{E}_{\mathbf{x}_{1:n}^{(0)}}\left[\left|\omega_n\left(\widehat{\pi}_{n-1}^{(m)}\right)\left(f_n\right)-\omega_n\left(\pi_{n-1}\right)\left(f_n\right)\right|^p\right]^{1/p} \\
&\leq 2D_n\mathbb{E}_{\mathbf{x}_{1:n}^{(0)}}\left[\left|\widehat{\pi}_{n-1}^{(m)}\left(\frac{\pi_{n/n-1}}{D_n}\right)-\pi_{n-1}\left(\frac{\pi_{n/n-1}}{D_n}\right)\right|^p\right]^{1/p} \\
&\leq\frac{2D_nC_{1,n-1}C_{2,p}}{(m+1)^{1/2}}
\end{aligned}$$

and

$$\mathbb{E}_{\mathbf{x}_{1:n}^{(0)}}\left[\left|S_n^{(i)}\left(f_n\right)-\pi_n\left(f_n\right)\right|^p\right]^{1/p}\leq\frac{2D_nC_{1,n-1}C_{2,p}}{(i+1)}\sum_{m=0}^{i}\frac{1}{(m+1)^{1/2}}\leq\frac{D_nC_{1,n-1}C_{2,p}}{(i+1)^{1/2}}.$$

This concludes the proof.■

A.4. Convergence of Marginals. **Proof of Theorem 3.2**. For $n=1$ the result follows directly from Assumption A1. Now consider the case where $n\geq 2$. We use the following decomposition for $0\leq n(i)\leq i$

$$\begin{aligned}
\left|\mathbb{E}_{\mathbf{x}_{1:n}^{(0)}}\left[f_n\left(\mathbf{X}_n^{(i)}\right)-\pi_n\left(f_n\right)\right]\right| &\leq\left|\mathbb{E}_{\mathbf{x}_{1:n}^{(0)}}\left[f_n\left(\mathbf{X}_n^{(i)}\right)-K_{n,\widehat{\pi}_{n-1}^{(i-n(i))}}^{n(i)}f_n\left(\mathbf{X}_n^{(i-n(i))}\right)\right]\right| \\
&+\left|\mathbb{E}_{\mathbf{x}_{1:n}^{(0)}}\left[K_{n,\widehat{\pi}_{n-1}^{(i-n(i))}}^{n(i)}\left(f_n\right)\left(\mathbf{X}_n^{(i-n(i))}\right)-\omega_n\left(\widehat{\pi}_{n-1}^{(i-n(i))}\right)\left(f_n\right)\right]\right| \\
&+\left|\mathbb{E}_{\mathbf{x}_{1:n}^{(0)}}\left[\omega_n\left(\widehat{\pi}_{n-1}^{(i-n(i))}\right)\left(f_n\right)-\omega_n\left(\pi_{n-1}\right)\left(f_n\right)\right]\right|
\end{aligned}$$

Assumption A1 implies that

$$\left|\mathbb{E}_{\mathbf{x}_{1:n}^{(0)}}\left[K_{n,\widehat{\pi}_{n-1}^{(i-n(i))}}^{n(i)}\left(f_n\right)\left(\mathbf{X}_n^{(i-n(i))}\right)-\omega_n\left(\widehat{\pi}_{n-1}^{(i-n(i))}\right)\left(f_n\right)\right]\right|\leq\rho_n^{n(i)}.$$

For the first term, we use the following decomposition

$$\begin{aligned}
&\mathbb{E}_{\mathbf{x}^{(0)}_{1:n}}\left[f_n\left(\mathbf{X}^{(i)}_n\right) - K^{n(i)}_{n,\widehat{\pi}^{(i-n(i))}_{n-1}}(f_n)\left(\mathbf{X}^{(i-n(i))}_n\right)\right] \\
&= \sum_{j=2}^{n(i)} \mathbb{E}_{\mathbf{x}^{(0)}_{1:n}}\left[K^{j-1}_{n,\widehat{\pi}^{(i-j+1)}_{n-1}}(f_n)\left(\mathbf{X}^{(i-j+1)}_n\right) - K^{j}_{n,\widehat{\pi}^{(i-j)}_{n-1}}(f_n)\left(\mathbf{X}^{(i-j)}_n\right)\right]
\end{aligned}$$

and

$$\begin{aligned}
&\mathbb{E}_{\mathbf{x}^{(0)}_{1:n}}\left[K^{j-1}_{n,\widehat{\pi}^{(i-j+1)}_{n-1}}(f_n)\left(\mathbf{X}^{(i-j+1)}_n\right) - K^{j}_{n,\widehat{\pi}^{(i-j)}_{n-1}}(f_n)\left(\mathbf{X}^{(i-j)}_n\right)\right] \\
&= \mathbb{E}_{\mathbf{x}^{(0)}_{1:n}}\left[\mathbb{E}_{\mathbf{x}^{(0)}_{1:n}}\left[K^{j-1}_{n,\widehat{\pi}^{(i-j+1)}_{n-1}}(f_n)\left(\mathbf{X}^{(i-j+1)}_n\right) - K^{j-1}_{n,\widehat{\pi}^{(i-j)}_{n-1}}(f_n)\left(\mathbf{X}^{(i-j+1)}_n\right)\middle|\,\mathcal{G}^{i-j}_n\right]\right]
\end{aligned}$$

where

$$\begin{aligned}
&K^{j-1}_{n,\widehat{\pi}^{(i-j+1)}_{n-1}}(f_n)\left(\mathbf{X}^{(i-j+1)}_n\right) - K^{j-1}_{n,\widehat{\pi}^{(i-j)}_{n-1}}(f_n)\left(\mathbf{X}^{(i-j+1)}_n\right) \\
&= \sum_{m=0}^{j-2} K^{m}_{n,\widehat{\pi}^{(i-j+1)}_{n-1}}\left(K_{n,\widehat{\pi}^{(i-j+1)}_{n-1}} - K_{n,\widehat{\pi}^{(i-j)}_{n-1}}\right)K^{j-1-m-1}_{n,\widehat{\pi}^{(i-j)}_{n-1}}(f_n)\left(\mathbf{X}^{(i-j+1)}_n\right) \\
&= \sum_{m=0}^{j-2} K^{m}_{n,\widehat{\pi}^{(i-j+1)}_{n-1}}\left(K_{n,\widehat{\pi}^{(i-j+1)}_{n-1}} - K_{n,\widehat{\pi}^{(i-j)}_{n-1}}\right) \\
&\times\left(K^{j-1-m-1}_{n,\widehat{\pi}^{(i-j)}_{n-1}}(f_n)\left(\mathbf{X}^{(i-j+1)}_n\right) - \omega_n\left(\widehat{\pi}^{(i-j)}_{n-1}\right)(f_n)\right).
\end{aligned}$$

Now we have from Proposition 3 that

$$\begin{aligned}
\left\|K^{m}_{n,\widehat{\pi}^{(i-j+1)}_{n-1}}(\mathbf{x}_n,\cdot) - K^{m}_{n,\widehat{\pi}^{(i-j)}_{n-1}}(\mathbf{x}_n,\cdot)\right\|_{\mathrm{tv}} &\leq \frac{2}{(1-\rho_n)}\left\|\widehat{\pi}^{(i-j+1)}_{n-1} - \widehat{\pi}^{(i-j)}_{n-1}\right\|_{\mathrm{tv}} \\
&\leq \frac{2}{(1-\rho_n)}\frac{1}{i-j+2}
\end{aligned}$$

and using A1

$$\begin{aligned}
&\left|\mathbb{E}_{\mathbf{x}_{1:n}^{(0)}}\left[\mathbb{E}_{\mathbf{x}_{1:n}^{(0)}}\left[K^{j-1}_{n,\widehat{\pi}_{n-1}^{(i-j+1)}}\left(f_n\right)\left(\mathbf{X}_n^{(i-j+1)}\right)-K^{j-1}_{n,\widehat{\pi}_{n-1}^{(i-j)}}\left(f_n\right)\left(\mathbf{X}_n^{(i-j+1)}\right)\middle|\mathcal{G}_n^{i-j}\right]\right]\right|\\
&\leq\left|\mathbb{E}_{\mathbf{x}_{1:n}^{(0)}}\left[\mathbb{E}_{\mathbf{x}_{1:n}^{(0)}}\left[\sum_{m=0}^{j-2}K^{m}_{n,\widehat{\pi}_{n-1}^{(i-j+1)}}\left(K_{n,\widehat{\pi}_{n-1}^{(i-j+1)}}-K_{n,\widehat{\pi}_{n-1}^{(i-j)}}\right)\right.\right.\right.\\
&\times\left.\left.\left.\left(K^{j-1-m-1}_{n,\widehat{\pi}_{n-1}^{(i-j)}}\left(f_n\right)\left(\mathbf{X}_n^{(i-j+1)}\right)-\omega_n\left(\widehat{\pi}_{n-1}^{(i-j)}\right)\left(f_n\right)\right)\middle|\mathcal{G}_n^{i-j}\right]\right]\right|\\
&\leq\frac{2}{\left(1-\rho_n\right)\left(i-j+2\right)}\sum_{m=0}^{j-2}\rho_n^{j-m-2}\\
&=\frac{2}{\left(1-\rho_n\right)\left(i-j+2\right)}\frac{1-\rho_n^{j-1}}{1-\rho_n}
\end{aligned}$$

and

$$\begin{aligned}
&\left|\mathbb{E}_{\mathbf{x}_{1:n}^{(0)}}\left[f_n\left(\mathbf{X}_n^{(i)}\right)-K^{n(i)}_{n,\widehat{\pi}_{n-1}^{(i-n(i))}}\left(f_n\right)\left(\mathbf{X}_n^{(i-n(i))}\right)\right]\right|\\
&\leq\frac{2}{\left(1-\rho_n\right)^2}\sum_{j=2}^{n(i)}\frac{1}{\left(i-j+2\right)}\\
&\leq\frac{2}{\left(1-\rho_n\right)^2}\log\left(\frac{i}{i-n\left(i\right)+1}\right).
\end{aligned}$$

Finally to study the last term $\mathbb{E}\left[\omega_n\left(\widehat{\pi}_{n-1}^{(i-n(i))}\right)\left(f_n\right)-\omega_n\left(\pi_{n-1}\right)\left(f_n\right)\right]$, we use the same decomposition used in the proof of Theorem 3.1 to obtain

$$\begin{aligned}
&\left|\mathbb{E}\left[\omega_n\left(\widehat{\pi}_{n-1}^{(i-n(i))}\right)\left(f_n\right)-\omega_n\left(\pi_{n-1}\right)\left(f_n\right)\right]\right|\\
&\leq 2D_n\mathbb{E}_{\mathbf{x}_{1:n}^{(0)}}\left[\left|\widehat{\pi}_{n-1}^{(i-n(i))}\left(\frac{\pi_{n/n-1}}{D_n}\right)-\pi_{n-1}\left(\frac{\pi_{n/n-1}}{D_n}\right)\right|\right]\\
&\leq\frac{2D_nC_{1,n-1}}{\left(i-n\left(i\right)+1\right)^{1/2}}.
\end{aligned}$$

One can check that $\left|\mathbb{E}_{\mathbf{x}_{1:n}^{(0)}}\left[f_n\left(\mathbf{X}_n^{(i)}\right)-\pi_n\left(f_n\right)\right]\right|$ converges towards zero for $n\left(i\right)=\lfloor i^{\alpha}\rfloor$ where $0<\alpha<1$.■

REFERENCES

[1] Barndorff-Nielsen, O. and Shephard, N. (2001) Non-Gaussian Ornstein-Uhlenbeck-based models and some of the uses in financial economics. *J. Royal Statist. Soc.* B, **63**, 167–241.

[2] Brockwell, A., Rojas, A. and Kass, R. (2004) Recursive Bayesian decoding of motor cortical signals by particle filtering. *J. Neurophysiology*, **91**, 1899-1907.
[3] Carpenter, J., Clifford, P. and Fearnhead, P. (1999) An improved particle filter for non-linear problems. *IEE Proceedings - Radar, Sonar and Navigation*, **146**, 2-7.
[4] Chib, S., Pitt, M.K. and Shephard, N. (2006) Likelihood based inference for diffusion driven models. Technical report, Nuffield College, Oxford.
[5] Chopin, N. (2002). A sequential particle filter method for static models. *Biometrika*, **89**, 539-552.
[6] Del Moral, P. (2004) *Feynman-Kac Formulae: Genealogical and Interacting Particle Systems with Applications*, Series Probability and Applications, New York: Springer-Verlag.
[7] Del Moral, P., Doucet, A. and Jasra, A. (2006) Sequential Monte Carlo samplers, *J. Royal Statist. Soc.* B, **68**, 411-436.
[8] Del Moral, P. and Miclo, L. (2003) On convergence of chains with occupational self-interactions, *Proc. Royal Soc. Lond.* A., **460**, 325–346.
[9] Doucet, A., Godsill, S.J. and Andrieu, C. (2000) On sequential Monte Carlo sampling methods for Bayesian filtering. *Statist. Comput.*, **10**, 197-208.
[10] Doucet, A., de Freitas, J.F.G. and Gordon, N.J. (eds.) (2001) *Sequential Monte Carlo Methods in Practice*. New York: Springer-Verlag.
[11] Gerlach, R., Carter, C.K. and Kohn, R. (1999) Diagnostics for time series analysis. *J. Time Ser. Analysis*, **20**, 309-330.
[12] Green, P.J. (2003) Trans-dimensional Markov chain Monte carlo. in *Highly Structured Stochastic Systems* (eds P.J. Green, N.L. Hjort and S. Richardson), 179-196, Oxford: Oxford University Press.
[13] Kitagawa, G. (1996) Monte Carlo filter and smoother for non-Gaussian nonlinear state space models. *J. Comp. Graph. Statist.*, **5**, 1-25.
[14] Kou, S., Zhou, Q. and Wong, W. (2006). Equi-energy sampler with applications in statistical inference and statistical mechanics (with discussion). *Ann. Statist.*, 34, 1581-1652.
[15] Jasra, A., Stephens, D.A. and Holmes, C.C. (2007) On population-based simulation for static inference. *Statist. Comput.*, **17**, 263-279.
[16] Liu, J.S. (2001) *Monte Carlo Strategies in Scientific Computing*. New York: Springer Verlag.
[17] Liu, J.S. and Chen, R. (1998) Sequential Monte Carlo for dynamic systems. *J. Amer. Statist. Assoc.*, **93**, 1032-44.
[18] Lyman, E. and Zuckerman, D.M. (2006) Resolution exchange simulation with incremental coarsening. *J. Chem. Theory Comp.*, **2**, 656-666.
[19] Mengersen, K.L. and Tweedie, R. (1996) Rates of convergence of the Hastings and Metropolis algorithms, *Ann. Statist.*, **24**, 101-121.
[20] Pitt, M.K. and Shephard, N. (1999). Filtering via simulation: auxiliary particle filter. *J. Am. Statist. Ass.*, **94**, 590-599.
[21] Shiryaev, A.N. (1996) *Probability*, 2nd edition, Springer-Verlag.
[22] Tierney, L. (1994) Markov chains for exploring posterior distributions. *Ann. Statist.*, **22**, 1701-1728.

Anthony Brockwell
Department of Statistics,
Carnegie Mellon University,
5000 Forbes Avenue,
Pittsburgh, PA 15123, USA.
E-mail: Abrock@stat.cmu.edu

Pierre Del Moral
Centre INRIA Bordeaux Sud-Ouest,
Institut de Mathématiques,
Université Bordeaux I,
33405 Talence cedex, France.
E-mail: Pierre.Del-Moral@inria.fr

Arnaud Doucet
Department of Statistics,
University of British Columbia,
333-6356 Agricultural Road,
Vancouver, BC, V6T 1Z2, Canada.
E-mail: Arnaud@stat.ubc.ca