

# Approximate Bayesian Computation for Smoothing

BY JAMES S. MARTIN<sup>1</sup>, AJAY JASRA<sup>2</sup>, SUMEETPAL S. SINGH<sup>3</sup>, NICK WHITELEY<sup>4</sup> & EMMA McCOY<sup>5</sup>

<sup>1</sup>Australian School of Business, University of New South Wales, Sydney, 2052, AUS.

E-Mail: *james.martin04@ic.ac.uk*

<sup>2</sup>Department of Statistics & Applied Probability, National University of Singapore, Singapore, 117546, SG.

E-Mail: *staja@nus.edu.sg*

<sup>3</sup>Department of Engineering, University of Cambridge, Cambridge, CB2 1PZ, UK.

E-Mail: *sss40@cam.ac.uk*

<sup>4</sup>Department of Mathematics, University of Bristol, Bristol, BS8 1TW, UK.

E-Mail: *nick.whiteley@bristol.ac.uk*

<sup>5</sup>Department of Mathematics, Imperial College London, London, SW7 2AZ, UK.

E-Mail: *e.mccoy@ic.ac.uk*

## Abstract

We consider a method for approximate inference in hidden Markov models (HMMs). The method circumvents the need to evaluate conditional densities of observations given the hidden states. It may be considered an instance of Approximate Bayesian Computation (ABC) and it involves the introduction of auxiliary variables valued in the same space as the observations. The quality of the approximation may be controlled to arbitrary precision through a parameter  $\epsilon > 0$ . We provide theoretical results which quantify, in terms of  $\epsilon$ , the ABC error in approximation of expectations of additive functionals with respect to the smoothing distributions. Under regularity assumptions, this error is  $\mathcal{O}(n\epsilon)$ , where  $n$  is the number of time steps over which smoothing is performed. For numerical implementation we adopt the forward-only sequential Monte Carlo (SMC) scheme of [16] and quantify the combined error from the ABC and SMC approximations. This forms some of the first quantitative results for ABC methods which jointly treat the ABC and simulation errors, with a finite number of data and simulated samples. When the HMM has unknown static parameters, we consider particle Markov chain Monte Carlo [2] (PMCMC) methods for batch statistical inference.

**Key-words:** Smoothing, Hidden Markov Models, Approximate Bayesian Computation, Sequential Monte Carlo, Markov chain Monte Carlo.

## 1 Introduction

Hidden Markov Models are widely used in statistics; see [9] for a recent overview. An HMM is a pair of discrete-time stochastic processes,  $\{X_n\}_{n \geq 0}$  and  $\{Y_n\}_{n \geq 0}$ , where  $X_n \in \mathbb{R}^{d_x}$  is unobserved and  $Y_n \in \mathbb{R}^{d_y}$  is observed. The hidden process  $\{X_n\}_{n \geq 0}$  is a Markov chain with initial density  $\eta_0$  at time 0 and transition density  $f(x_{n-1}, x_n)$ , i.e.

$$\mathbb{P}(X_0 \in A) = \int_A \eta_0(x) dx \quad \text{and} \quad \mathbb{P}(X_n \in A | X_{n-1} = x_{n-1}) = \int_A f(x_{n-1}, x_n) dx_n \quad n \geq 1 \quad (1)$$

where  $A \subseteq \mathbb{R}^{d_x}$  and  $dx_n$  is a dominating  $\sigma$ -finite measure. Each observation  $Y_n$  is conditionally independent of other variables given  $\{X_n\}_{n \geq 0}$  and its conditional distribution is specified by a density  $g(x_n, y_n)$ , i.e.

$$\mathbb{P}(Y_n \in B | \{X_k\}_{k \geq 0} = \{x_k\}_{k \geq 0}) = \int_B g(x_n, y_n) dy_n \quad n \geq 0 \quad (2)$$

with  $B \subseteq \mathbb{R}^{d_y}$  and  $dy_n$  is a dominating  $\sigma$ -finite measure. We remark that  $\eta_0$ ,  $f(x_{n-1}, x_n)$  and  $g(x_n, y_n)$  may depend upon time-independent parameters, which we term static parameters.

A variety of inference and estimation tasks for HMMs involve the computation of the smoothing functional  $\mathcal{V}_n : \mathbb{R}^{(n+1)d_x} \rightarrow \mathbb{R}$

$$\mathbb{E}[\mathcal{V}_n(X_{0:n}) | y_{0:n}] \quad (3)$$

where  $\mathcal{V}_n(x_{0:n}) = \sum_{p=0}^n v_p(x_{p-1:p})$ ,  $v_p : \mathbb{R}^{2d_x} \rightarrow \mathbb{R}$ ,  $1 \leq p \leq n$ ,  $x_{-1:0} = x_0$  with  $v_0 : \mathbb{R}^{d_x} \rightarrow \mathbb{R}$  and the expectation is w.r.t. the joint smoothing distribution. For example in the cases  $v_p(x_p) = x_p/(n+1)$  and  $v_p(x_{p-1:p}) = x_{p-1}x_p/(n+1)$ , (3) approximates the posterior mean and first-order auto-covariance. When the HMM includes unknown static parameters, expectations of additive functionals play a central role in expectation-maximisation (EM) algorithms and the calculation of score vectors; see [9, 16] for some discussion.

In practice, the expectation in (3) can rarely be computed exactly and one resorts to the use of numerical integration techniques such as SMC; see [18] for a recent overview. These methods typically rely on the ability to evaluate pointwise the conditional density  $g(x, y)$ . The methods we consider address the problem of obtaining an approximation of (3) without performing any such evaluations, in a principled manner admitting control of the error in approximation. The motivations for avoiding such evaluations are as follows. Firstly, for some models,  $g(x, y)$  may simply not have a closed form expression. Secondly, in some situations evaluation of  $g(x, y)$  may be very expensive.

The general technique we consider may be interpreted as an instance of ABC. A recent review of this class of methods can be found in [24]. In the context of HMMs, ABC has been considered by [12, 22], see also [11]. The approximation (given in Section 2.1) has been introduced by [25] and still requires numerical (e.g. SMC) methods to fit them. Alternative ideas include nonparametric filtering [20] and the related convolution particle filter [8]; see [22] for some discussion relative to ABC. As noted by [24], in the scenario where there is a fixed amount of data and a fixed number of simulated samples, there is a distinct lack of theoretical results which quantify the combined ABC and numerical (SMC) errors; we provide some of the first results in this context (see [7, 22] for other results).

Our main objectives are to

1. investigate, both theoretically and empirically, the error associated with the approximation scheme we propose
2. demonstrate how this scheme can be used to perform smoothing and to estimate static parameters in HMMs from a batch of data

Regarding point 1., the error has two components. The first component arises from the introduction of an auxiliary HMM, incorporating auxiliary variables valued in the same space as the observations  $\{y_n\}_{n \geq 0}$ . Smoothing expectations under this auxiliary model, which we write as  $\mathbb{E}^\epsilon[\mathcal{V}_n(X_{0:n})|y_{0:n}]$ , may be taken as approximations of (3); the degree of approximation is controlled through a parameter  $\epsilon$  and the error should disappear as  $\epsilon \rightarrow 0$ . In turn, expectations under this auxiliary model admit efficient numerical approximation using SMC techniques, without evaluation of  $g(x, y)$ . The second component of the overall error arises from this Monte Carlo scheme, and is controlled through a sample size parameter  $N$ ; the error disappears as  $N \rightarrow +\infty$ . It is noted that the SMC method adopted is the forward only smoothing implementation of the forward-filtering backward smoothing (FFBS) method [3, 21] in [15]; this is currently one of the most accurate methods for SMC approximation of smoothed additive functionals.

We will write the SMC estimate of the smoothing expectation under the auxiliary HMM as  $\Xi_n[\epsilon, N, \mathcal{V}_n, y_{0:n}]$  and, similarly, we may denote the SMC estimate of the smoothing expectation under the original HMM (1)-(2) as  $\Xi_n[N, \mathcal{V}_n, y_{0:n}]$ . In the numerical studies in Section 5, the errors associated with these SMC estimates will be denoted as  $e_n^{N, \epsilon}$  and  $e_n^N$ , respectively.

The overall error associated with the SMC estimate of the smoothing expectation under the auxiliary HMM may be decomposed as

$$\begin{aligned} \mathbb{E}[\mathcal{V}_n(X_{0:n})|y_{0:n}] - \Xi_n[\epsilon, N, \mathcal{V}_n, y_{0:n}] &= \mathbb{E}[\mathcal{V}_n(X_{0:n})|y_{0:n}] - \mathbb{E}^\epsilon[\mathcal{V}_n(X_{0:n})|y_{0:n}] \\ &\quad + \mathbb{E}^\epsilon[\mathcal{V}_n(X_{0:n})|y_{0:n}] - \Xi_n[\epsilon, N, \mathcal{V}_n, y_{0:n}]. \end{aligned} \quad (4)$$

The first difference on the right of (4) is a deterministic error, the second difference is a stochastic error. We will provide theoretical analysis of these two error terms which shows how the interplay between  $\epsilon$ ,  $n$  and  $N$  controls the overall quality of the approximation. These theoretical results, to an extent, are also studied from an empirical perspective.

Regarding point 2., we show how the approximation scheme can be incorporated into a particle MCMC scheme in order to estimate static parameters of the HMM. Particle MCMC uses SMC techniques to generate proposals, for example, associated to hidden states of the HMM. Here, we use the particle marginal Metropolis-Hastings algorithm in [2] to sample from the ABC approximation of the HMM, with a prior placed upon the unknown static-parameters. In contexts where additive functions of the hidden state are also of interest (as above), we use the forward-only smoothing technique mentioned above, to use all the simulated samples from the SMC proposal (which is not typically adopted). A similar idea has been adopted by [26], except using an additional ‘backward-pass’ in the FFBS algorithm, which is not needed.

In summary, our main contributions are to:

- quantify, in terms of  $\epsilon$  and  $n$ , the error  $\mathbb{E}[\mathcal{V}_n(X_{0:n})|y_{0:n}] - \mathbb{E}^\epsilon[\mathcal{V}_n(X_{0:n})|y_{0:n}]$  - henceforth referred to as the *ABC error*

- quantify, in terms of  $\epsilon$ ,  $N$  and  $n$ , the error  $\mathbb{E}^\epsilon[\mathcal{V}_n(X_{0:n})|y_{0:n}] - \Xi_n[\epsilon, N, \mathcal{V}_n, y_{0:n}]$  - henceforth referred to as the *SMC error*
- provide empirical evidence which illustrates some of these theoretical findings

This paper is structured as follows. In Section 2 we discuss the ABC approximation and we characterize the ABC error. In Section 3, SMC and MCMC simulation techniques for targeting the (ABC) smoother are detailed, and the estimation of static parameters in a Bayesian manner is also considered. In Section 4 our main theoretical result is given, which combines the ABC and SMC errors discussed above. In Section 5 some numerical studies are presented. In Section 6 the article is concluded. The proofs are in the appendices.

## 1.1 Notations

Given a measurable space  $(E, \mathcal{E})$ , let  $\mu$  be a  $\sigma$ -finite measure,  $K$  be a non-negative kernel and  $f : E \rightarrow \mathbb{R}$  a measurable function. The conventions  $\mu(f) := \int_E f(x)\mu(dx)$ ,  $K(f)(x) := \int_E K(x, dy)f(y)$ ,  $\mu K(f) := \mu(K(f))$  are used. In addition, let  $\text{Osc}(f) := \sup_{(x,y) \in E^2} |f(x) - f(y)|$  and let  $\mathcal{B}_b(E)$  be the Banach space of bounded and measurable functions on  $E$  endowed with the norm  $\|f\| := \sup_{x \in E} |f(x)|$ . For two probability measures  $\mu_1, \mu_2$  the total variation distance is  $\|\mu_1 - \mu_2\|_{TV} = \sup_{A \in \mathcal{E}} |\mu_1(A) - \mu_2(A)|$ . For a Markov kernel  $K$  the Dobrushin coefficient is  $\beta(K) := \sup_{(x,y) \in E^2} \|K(x, \cdot) - K(y, \cdot)\|_{TV}$ . Given a probability space  $(\Omega, \mathcal{F}, \mathbb{P})$ , we write  $\|\cdot\|_p = \mathbb{E}[|\cdot|^p]^{1/p}$  for the  $\mathbb{L}_p$  norm under  $\mathbb{P}$ . For a set  $A \in \mathcal{E}$ ,  $\mathbb{I}_A(x)$  is the indicator function.

## 2 ABC Approximation

### 2.1 ABC Smoothing Approximation

The joint smoothing density is

$$\hat{\eta}_n(x_{0:n}) = \frac{\prod_{i=0}^n g(x_i, y_i) \eta_0(x_0) \prod_{i=1}^n f(x_{i-1}, x_i)}{\int_{\mathbb{R}^{(n+1)d_x}} \prod_{i=0}^n g(x_i, y_i) \eta_0(x_0) \prod_{i=1}^n f(x_{i-1}, x_i) dx_{0:n}}$$

where we will suppress the dependence on the data on the l.h.s. In most scenarios of practical interest, one cannot calculate this density pointwise, or compute expectations w.r.t. the density. As a result, a numerical approximation of (3), via advanced computational tools, is required. This problem is further exacerbated when the density  $g(x, y)$  is intractable or very expensive to calculate. That is, one cannot evaluate it pointwise and there is no unbiased estimate available. However, we will assume throughout that one can sample from the associated distribution, for any  $x \in \mathbb{R}^{d_x}$ . It is remarked that this latter condition is not completely necessary for any of the subsequent ideas that will appear (see [19, 28]), but, it will facilitate a more compact exposition.

To introduce ideas, let us momentarily step away from the setting of HMM's; suppose, one is given observations  $y \in \mathcal{D}$  associated to some intractable likelihood  $g_\theta(y)$  with  $\theta \in \Theta$  an unknown parameter. Then, Bayesian inference associated to the posterior  $\pi(\theta|y) \propto g_\theta(y)\pi(\theta)$  is typically not feasible even using advanced computational tools; see [24]. To deal with this issue, ABC draws inference from the following modified posterior density on  $\Theta \times \mathcal{D}$

$$\pi_\epsilon(\theta, u|y) = \frac{\pi(\theta)g_\theta(u)\mathbb{I}_{A_{\epsilon,y}}(u)}{\int_{A_{\epsilon,y} \times \Theta} \pi(\theta)g_\theta(u)dud\theta} \quad (5)$$

with  $\epsilon > 0$  a tolerance level and  $u \in \mathcal{D}$  corresponds to some pseudo-observations. The set  $A_{\epsilon,y}$  is defined as follows

$$A_{\epsilon,y} = \{z \in \mathcal{D} : \rho(s(z), s(y)) < \epsilon\}$$

where  $s : \mathcal{D} \rightarrow \mathcal{S}$  represents some summary statistics and  $\rho : \mathcal{S} \times \mathcal{S} \rightarrow \mathbb{R}^+$  a distance metric.

As noted by [4, 22], given an appropriate structure for the likelihood (such as i.i.d. data) one can often achieve a more accurate approximation by removing the summary statistics and focusing upon the probabilistic structure of the likelihood. Returning now to our setting of the HMM specified in section 1, following [22, 25] we consider the ABC approximation of the joint smoothing density, for  $\epsilon > 0$ :

$$\hat{\eta}_{n,\epsilon}(x_{0:n}) = \frac{\prod_{i=0}^n \int_{\mathbb{R}^{d_y}} \phi\left(\frac{u-y_i}{\epsilon}\right) g(x_i, u) du \eta_0(x_0) \prod_{i=1}^n f(x_{i-1}, x_i)}{\int_{\mathbb{R}^{(n+1)d_x}} \prod_{i=0}^n \int_{\mathbb{R}^{d_y}} \phi\left(\frac{u-y_i}{\epsilon}\right) g(x_i, u) du \eta_0(x_0) \prod_{i=1}^n f(x_{i-1}, x_i) dx_{0:n}} \quad (6)$$

where  $\phi\left(\frac{u-y}{\epsilon}\right)$  is a potential which can take the form  $\mathbb{1}_{\{|u-(u-y)/\epsilon|<1\}}$  or a probability density function. This particular ABC approximation maintains the Markovian structure of the model, which will help to facilitate computational algorithms. In particular, in order to approximate  $\mathbb{E}^\epsilon[\mathcal{V}_n(X_{0:n})|y_{0:n}]$  where  $\mathbb{E}^\epsilon[\cdot|y_{0:n}]$  is an expectation w.r.t. the ABC smoothing distribution, one will still need to resort to numerical methods such as SMC and MCMC.

From the perspective of a theoretical justification, results on the consistency in estimation (from both classical and Bayesian perspectives) of static parameters, associated to the ABC approximation, as  $n$  grows can be found in [12, 13]. There is an intrinsic asymptotic bias, but this bias can be removed by using a noisy version of ABC; see [12, 13] for further details. In this article we do not address the idea of noisy ABC.

## 2.2 ABC Error

In order to ascertain the potential of using an ABC approximation of HMMs, when considering smoothed additive functionals, we investigate the ABC error. The analysis that follows will concentrate on the scenario in which the ABC kernel is *not* the indicator function. In particular, this allows the inclusion of (A1) (3), below, which will facilitate the theoretical analysis of the SMC error of the ABC smoother, in Section 4. We remark that the below result for the ABC bias can also be established when  $\phi\left(\frac{u-y}{\epsilon}\right)$  is an indicator function with some minor modifications to the proof.

In the subsequent analysis we adopt the following assumption.

- (A1) 1. There exists a  $1 < \rho < \infty$  such that for each  $x, x' \in \mathbb{R}^{d_x}$  and  $y \in \mathbb{R}^{d_y}$

$$\begin{aligned} \rho^{-1} &\leq g(x, y) \leq \rho \\ \rho^{-1} &\leq f(x, x') \leq \rho. \end{aligned}$$

2. There exists a  $L < \infty$  such that for every  $y, y' \in \mathbb{R}^{d_y}$

$$\sup_{x \in \mathbb{R}^{d_x}} |g(x, y) - g(x, y')| \leq L|y - y'|$$

with  $|\cdot|$  the  $L_1$ -norm.

3. There exist functions  $\bar{\alpha}, \underline{\alpha} : \mathbb{R}_+ \rightarrow \mathbb{R}_+$  such that for any  $y, u \in \mathbb{R}^{d_y}$ ,  $\epsilon \in \mathbb{R}_+$

$$\underline{\alpha}(\epsilon) \leq \phi\left(\frac{y-u}{\epsilon}\right) \leq \bar{\alpha}(\epsilon)$$

with  $\delta(\epsilon) := \bar{\alpha}(\epsilon)/\underline{\alpha}(\epsilon)$  monotonically decreasing. In addition  $\int_{\mathbb{R}^{d_y}} \phi\left(\frac{y-u}{\epsilon}\right) dy = 1$  with  $\int_{\mathbb{R}^{d_y}} |y| \phi(y) dy < +\infty$ .

4. The analysis is associated to additive functionals  $V_n : \mathbb{R}^{(n+1)d_x} \rightarrow \mathbb{R}$ :

$$\mathcal{V}_n(x_{0:n}) = \sum_{p=0}^n v_p(x_p).$$

with  $\bar{v} = \sup_{0 \leq p < \infty} \|v_p\| < +\infty$ .

The assumptions are rather strong and will typically only hold when the observations and hidden states lie on a compact state spaces, they are however, quite typically of assumptions employed in the analysis of SMC and related approximation methods. Of these assumptions, perhaps (A1-3) should be discussed. It can be verified when:

$$\phi\left(\frac{y-u}{\epsilon}\right) \propto \exp\left\{-\left(\frac{y-u}{\epsilon}\right)\right\} \quad y \in (a, b).$$

It is remarked that in practice, one selects  $\phi\left(\frac{y-u}{\epsilon}\right)$ , so this is not such a demanding assumption.

### 2.2.1 Result

We have the following result, whose proof is in Appendix B.

**Theorem 2.1.** *Assume (A1). Then there exist a  $C < +\infty$  such that for any  $\epsilon > 0$ ,  $n \geq 1$ ,  $y_{0:n}$ ,*

$$|\mathbb{E}[\mathcal{V}_n(X_{0:n})|y_{0:n}] - \mathbb{E}^\epsilon[\mathcal{V}_n(X_{0:n})|y_{0:n}]| \leq C\epsilon(n+1)$$

where  $\mathbb{E}[\cdot|y_{0:n}]$  and  $\mathbb{E}^\epsilon[\cdot|y_{0:n}]$  are the expectation w.r.t. the joint smoothing and ABC smoothing distribution.

**Remark 2.1.** *The result establishes that the ABC error does not grow any faster than linearly in time or  $\epsilon$ . This is important, as it is known that the SMC error when estimating  $\mathbb{E}^\epsilon[\mathcal{V}_n(X_{0:n})|y_{0:n}]$  also grows at most linearly in time [15]. As a result, the overall error as the time parameter increases will not necessarily be dominated by one source of error (SMC or ABC). This suggests that an ABC approximation can perform reasonably well in general.*

### 3 Simulation-Based Methods

#### 3.1 SMC Methods

In the context of HMMs, SMC algorithms approximate  $\{\hat{\eta}_n\}$  recursively by propagating a collection of properly weighted samples, called particles, using a combination of importance sampling and resampling steps. For the importance sampling part of the algorithm at each step  $n$  of the algorithm we will use general proposal (Markov) kernels  $H_n$  which possess normalizing constants that do not depend on the simulated paths. A typical SMC algorithm is given below (we assume it terminates at time  $p + 1$ ):

- 0. Initialisation: set  $n = 0$ ; for  $i \in \{1, \dots, N\}$  sample  $X_0^i \sim \eta_0$  and compute

$$G_0(x_0^i) = g(x_0^i, y_0)$$

with  $W_0^i = G_0(x_0^i)$ .

- 1. Decide whether or not to resample, and if this is performed, set all weights  $\{W_n^{(i)}\}_{1 \leq i \leq N}$  to 1. Proceed to step 2.
- 2. Set  $n = n + 1$ , if  $n = p + 1$  stop, else; for  $i \in \{1, \dots, N\}$  sample  $X_n^i | x_{n-1}^i \sim H_n(x_{n-1}^i, \cdot)$ , compute

$$G_n(x_{n-1:n}^i) = \frac{g(x_n^i, y_n) f(x_{n-1}^i, x_n^i)}{H_n(x_{n-1}^i, x_n^i)}$$

and set  $W_n^i = G_n(x_{n-1:n}^i) W_{n-1}^i$  and return to the start of step 1.

#### 3.2 Some details on resampling

If one chooses to implement SMC without resampling steps, i.e. to perform sequential importance sampling, as time progresses, the variance of the weights  $\{W_n^i\}_{1 \leq i \leq N}$  typically increases. This has been commonly referred to as the *weight* degeneracy property. To counter this resampling is used: the particles are sampled with replacement, according to the normalized weights  $\{\bar{W}_n^i\}_{1 \leq i \leq N}$  given by  $\bar{W}_n^i = \frac{W_n^i}{\sum_{j=1}^N W_n^j}$  and then each  $W_n^i$  is reset to 1. We remark that more efficient alternatives are possible; see e.g. [18].

If one resamples too often, the simulated past of the path of each particle will be very similar to each other. This has been documented as the *path* degeneracy problem. A common remedy was to resample only when an appropriate criterion drops beneath or goes above some threshold. In the former case, a common criterion is the effective sample size  $\left(\sum_{j=1}^N (\bar{W}_n^j)^2\right)^{-1}$  [23]. This approach, however, does not ultimately solve the path degeneracy problem. Path degeneracy has been a long standing bottleneck when static parameters  $\theta$  are estimated online using SMC methods by augmenting them with the latent state; see [16]. Considering the central limit theorem (CLT) associated to the SMC estimate of  $\mathbb{E}[\mathcal{V}_n(X_{0:n})|y_{0:n}]$ :

$$\sum_{i=1}^N \bar{W}_n^i \mathcal{V}_n(x_{0:n}^i),$$

it is remarked that the issue of path degeneracy leads, under very strong conditions on the HMM, to an asymptotic variance in this CLT that grows quadratically in  $n$ ; see [27].

Suppose one resamples, multinomially, at every iteration, except when  $n = p$ . Denote the resampled index of the ancestor of particle  $i$  at time  $n$  by  $a_n^i \in \{1, \dots, N\}$ ; this is a random variable chosen with probability  $\bar{W}_{n-1}^{a_n^i}$ . Furthermore the joint density of the sampled particles and the resampled indices is

$$\psi(x_{0:p}^{1:N}, \bar{\mathbf{a}}_{0:p-1}) = \left(\prod_{i=1}^N \eta_0(x_1^i)\right) \prod_{n=1}^p \left(\prod_{i=1}^N \bar{W}_{n-1}^{a_n^i} H_n(x_{n-1}^{a_n^i}, x_n^i)\right), \quad (7)$$

where the complete genealogy of ancestors is denoted as  $\bar{\mathbf{a}}_n = (a_n^1, \dots, a_n^N)$  and the randomly simulated values of the state as  $x_n = (x_n^1, \dots, x_n^N)$ . Together they form the following SMC approximations for  $\hat{\eta}_n$

$$\hat{\eta}_n^N(dx_{0:n}) = \frac{1}{N} \sum_{j=1}^N \delta_{x_{0:n}^{a_n^j}}(dx_{0:n})$$

and an approximation of the normalizing constant

$$\hat{Z}_p = \prod_{n=0}^p \left\{ \frac{1}{N} \sum_{j=1}^N G_n(x_{n-1:n}^j) \right\}. \quad (8)$$

The complete ancestral genealogy at each time can always be traced back by defining an ancestry sequence  $b_{0:n}^i$  for every  $i \in \{1, \dots, N\}$  and  $n \in \{0, \dots, p-1\}$ , whose elements are given by the backward recursion  $b_n^i = a_n^{b_{n-1}^i}$  where  $b_p^i = i$ . This interpretation of SMC approximations was introduced in [2] and will be used later together with  $\psi(x_{0:p}^{1:N}, \bar{\mathbf{a}}_{0:p-1})$  for describing PMCMC.

### 3.3 Forward only Smoothing

Due to the path degeneracy effect, one does not want to use the SMC approximation  $\hat{\eta}_n^N(dx_{0:n})$  to perform smoothing. One potential solution to this issue is the forward filtering backward smoothing algorithm and in particular the forward only implementation of it in [16]. That is, the FFBS algorithm includes a backward simulation step, which is eliminated in [16]. We consider the SMC approximation of the expectation  $\mathbb{E}[\mathcal{V}_n(X_{0:n})|y_{0:n}]$  where  $\mathcal{V}_n(x_{0:n}) = \sum_{p=0}^n v_p(x_{p-1:p})$ .

The construction of the procedure is as follows. It is first noted that

$$\mathbb{E}[\mathcal{V}_n(X_{0:n})|y_{0:n}] = \int V_n(x_n) \hat{\eta}_n(x_{0:n}) dx_{0:n}$$

with, for  $n \geq 1$

$$V_n(x_n) := \int \mathcal{V}_n(x_{0:n}) \hat{\eta}_n(x_{0:n-1}|x_n) dx_{0:n-1}$$

$V_0(x_0) = 0$ , where  $\hat{\eta}_n(x_{0:n-1}|x_n) = \hat{\eta}_n(x_{0:n}) / \int \hat{\eta}_n(x_{0:n}) dx_{0:n-1}$ , then one can establish, e.g. [16], that

$$V_n(x_n) = \int [V_{n-1}(x_{n-1}) + v_n(x_{n-1:n})] \hat{\eta}_n(x_{n-1}|x_n) dx_{n-1}$$

where  $\hat{\eta}_n(x_{n-1}|x_n) = \int \hat{\eta}_n(x_{0:n-1}|x_n) dx_{0:n-2}$ .

These recursions lead to the following idea. Given the current particle approximation of the marginal of  $\hat{\eta}_{n-1}$ ,  $\{\bar{W}_{n-1}^i, x_{n-1}^i\}_{1 \leq i \leq N}$  and of  $\{V_{n-1}(x_{n-1}^i)\}_{1 \leq i \leq N}$  (write this  $\{V_{n-1}^N(x_{n-1}^i)\}_{1 \leq i \leq N}$ ), one performs the following: update the SMC approximation as in Section 3.1 and set

$$V_n^N(x_n^i) = \frac{\sum_{j=1}^N \bar{W}_{n-1}^j f(x_{n-1}^j, x_n^i) [V_{n-1}^N(x_{n-1}^j) + v_n(x_{n-1}^j, x_n^i)]}{\sum_{j=1}^N \bar{W}_{n-1}^j f(x_{n-1}^j, x_n^i)} \quad i \in \{1, \dots, N\} \quad (9)$$

with  $V_0^N = v_0$ . Then the SMC approximation of  $\mathbb{E}[\mathcal{V}_n(X_{0:n})|y_{0:n}]$ , is exactly:

$$\sum_{i=1}^N \bar{W}_n^i V_n^N(x_n^i). \quad (10)$$

It is apparent that the computational cost of this recursion is  $\mathcal{O}(N^2)$  per-time step. For functions such as  $\mathcal{V}_n(x_{0:n}) = \sum_{p=0}^n v_p(x_{p-1:p})$ , it has been seen that, under some assumptions, the asymptotic variance in the CLT associated to (10) grows at most linearly in  $n$ . This is in contrast to growing quadratically at least quadratically in  $n$ , under similar assumptions, for the standard SMC estimate; see [15] and also [17] for additional theoretical analysis.

### 3.4 SMC for ABC

If one cannot or does not want to compute  $g(x, y)$ , then the algorithm described in Section 3.1 can seldom be implemented. In such contexts, we can easily use an SMC algorithm to approximate the  $\hat{\eta}_{n,\epsilon}$  in (6). For example, in [22], at time 0, one samples the signal from  $\eta_0$  and the pseudo observations from the likelihood to yield an incremental weight

$$G_0(u_0^i) = \phi\left(\frac{u_0^{(i)} - y_0}{\epsilon}\right)$$

where  $u_0^i$  is the pseudo observation at time 0. At subsequent time-points one can sample from the signal transition and likelihood to obtain  $G_n(u_n^i) = \phi\left(\frac{u_n^{(i)} - y_n}{\epsilon}\right)$ . The selection of  $\epsilon$  can be adaptive and different proposals (other than the state-dynamics) can be adopted; we refer to [22] for some discussion.

It is remarked that a drawback of the algorithm is that when  $d_y$  grows with  $\epsilon$ ,  $N$  fixed, one cannot expect the algorithm to work well for every  $\epsilon$ ; typically one must increase  $\epsilon$  to yield reasonable algorithmic results and this is at the cost of increasing the bias (see Theorem 2.1). To maintain  $\epsilon$  at a reasonable level, one must consider more advanced strategies which are not investigated here.

In scenarios where  $\phi\left(\frac{u-y}{\epsilon}\right) = \mathbb{I}_{\{|u-y|/\epsilon < 1\}}\left(\frac{u-y}{\epsilon}\right)$ , a potentially better procedure is to use the rejection kernel in [14] (note this differs from the ideas of [11]). In this case, one initializes the SMC algorithm as above. However, at subsequent time-points,  $n \geq 1$ , one uses the kernel

$$\begin{aligned} K_n((u_{n-1}^{1:N}, x_{n-1}^{1:N}), (u_n^i, x_n^i)) &= G_{n-1}(u_{n-1}^i) H_n((u_{n-1}^i, x_{n-1}^i), (u_n^i, x_n^i)) \\ &+ [1 - G_{n-1}(u_{n-1}^i)] \sum_{j=1}^N \frac{G_{n-1}(u_{n-1}^j)}{\sum_{l=1}^N G_{n-1}(u_{n-1}^l)} H_n((u_{n-1}^j, x_{n-1}^j), (u_n^i, x_n^i)). \end{aligned} \quad (11)$$

In this case, at any given time-step, we will only resample those particles which have  $|u_{n-1} - y_{n-1}| > \epsilon$ . It has been shown by [14, pp. 304-305] that this kernel produces a lower asymptotic variance in the CLT than an algorithm which resamples at every time step. It will be of interest to see if this advantage is realized when  $N$  is finite, especially versus the dynamic resampling that is mentioned in Section 3.1. This particular SMC approach is termed ‘rejection SMC’ (RSMC) throughout the article.

When using SMC for the ABC approximation of  $\mathbb{E}[\mathcal{V}_n(X_{0:n})|y_{0:n}]$ , the procedure in Section 3.3 can be followed with only modifications in notations and state-spaces.

### 3.5 PMCMC

In this section we consider the scenario where one has unknown static parameters  $\theta \in \mathbb{R}^{d_\theta}$  associated to the HMM. We concentrate upon batch inference.

Particle Markov Chain Monte Carlo methods are MCMC algorithms operating on an extended state-space and targeting an extended distribution over the random variables appearing in the SMC algorithm. As in standard MCMC the idea is to run an ergodic Markov chain to obtain samples from the distribution of interest. The difference lies in the fact that, due to using an SMC approximation to generate a proposal, the invariant distribution of the simulated chain is defined on an extended state space, with an appropriate marginal being the distribution that we are interested in sampling from in the first place.

We will present the particle marginal Metropolis-Hastings (PMMH) algorithm of [2]. The PMMH algorithm can sample from the target distribution

$$\hat{\eta}_n(x_{0:n}, \theta) \propto \left[ \prod_{i=0}^n g_\theta(x_i, y_i) \eta_0(x_0) \prod_{i=1}^n f_\theta(x_{i-1}, x_i) \right] \pi(\theta) \quad (12)$$

where  $\pi(\theta)$  is the prior on  $\theta$ . We concentrate upon the presentation in the scenario that one is interested in the original HMM; the ABC extension is simple and just uses the SMC procedures described in Section 3.4 instead of those at the start of Section 3.1. Note also, that the algorithm is given when using an SMC algorithm that resamples at each time-step; a dynamic resampling schedule can also be used.

The PMMH algorithm proceeds as follows:

- 0. Set  $\theta(0)$ . Sample  $x_{0:p}(0)^{1:N}$ ,  $\bar{\mathbf{a}}_{0:p-1}(0)$  from (7) (which now depends upon  $\theta$ ). Sample  $k \in \{1, \dots, N\}$  from  $\bar{W}_p^k$  and compute  $\hat{Z}_p^k(0)$  as in (8). Store  $\hat{Z}_p^k(0), k(0), x_{0:p}(0)^{1:N}, \bar{\mathbf{a}}_{1:p-1}(0), \theta(0)$ . Set  $i = 1$

- 1. Propose a new  $\theta'$  from a candidate  $q(\theta(i-1), \cdot)$  and  $x'_{0:p}, \bar{\mathbf{a}}'_1, \dots, \bar{\mathbf{a}}'_{p-1}$  and  $k'$  as in step 0. Accept or reject this as the new state of the chain with probability

$$1 \wedge \frac{\widehat{Z}'}{\widehat{Z}(i-1)} \frac{\pi(\theta')q(\theta', \theta(i-1))}{\pi(\theta(i-1))q(\theta(i-1), \theta')}$$

If we accept, set  $(\widehat{Z}(i), k(i), x_{0:p}^{1:N}(i), \bar{\mathbf{a}}_{1:p-1}(i), \theta(i)) = (\widehat{Z}', k', x'_{0:p}, \bar{\mathbf{a}}'_{1:p-1}, \theta')$ , otherwise  $(\widehat{Z}(i), k(i), x_{0:p}^{1:N}(i), \bar{\mathbf{a}}_{1:p-1}(i), \theta(i)) = (\widehat{Z}(i-1), k(i-1), x_{0:p}^{1:N}(i-1), \bar{\mathbf{a}}_{1:p-1}(i-1), \theta(i-1))$ . Set  $i = i + 1$  and return to 1.

In [2] it is shown that the sequence  $\{x_{0:p}^{b_{0:p}^{k(i)}}(i), \theta(i)\}_{i \geq 1}$  provides an approximation of (12), for any  $N \geq 1$ . If one is interested in approximating, say

$$\int \mathcal{V}_n(x_{0:n}, \theta) \hat{\eta}_n(x_{0:n}, \theta) d(x_{0:n}, \theta)$$

as noted by [26], the FFBS estimate can be used, based upon the SMC at each time-step, by simply extending the definition of  $V_n^N$  in (9) to include  $\theta$  (c.f. (10)):

$$\frac{1}{M} \sum_{i=1}^M \left\{ \sum_{j=1}^N \bar{W}_n^j(i) V_n^N(x_n^j(i), \theta(i)) \right\}$$

where the first summation is over  $M$  iterations of the PMMH algorithm. Trivially, one can extend this to the case where only a forward pass as in Section 3.3 is used. A critical point is despite the improvement in the SMC estimation, whether this is necessarily reasonable given the increase in computational cost and the iterative nature of the MCMC; especially in an ABC context, which is presently not known to our knowledge. We remark again, that any of the SMC for ABC algorithms mentioned in Section 3.4 can be adopted, when considering the ABC approximation of (12); whether using dynamic resampling or the kernel (11) the estimate of the normalizing constant is unbiased - see [2] for why this is of interest.

## 4 Theoretical Analysis

### 4.1 Set-Up

We consider the error in estimation of smoothed additive functionals, when using an ABC approximation of the HMM. This is in the scenario where one does not need to estimate static parameters. Recall that we have already considered the ABC error in Theorem 2.1; the main objective is to present a result with regards to the SMC error and the overall effect on the approximation of  $\mathbb{E}[\mathcal{V}_n(X_{0:n})|y_{0:n}]$ . We will use (A1) which only applies in the scenario where one uses a kernel density in the ABC approximation (i.e. not an indicator function). In addition, the SMC algorithm samples from the transition density of the state, with multinomial resampling at every time step (that is, RSMC is not considered). These hypotheses can be removed with a more technical proof. In addition, we condition upon the data and do not treat the randomness of these quantities. We simply assume that we are given a data set and do not address the issue of whether they may, or may not originate from a HMM.

### 4.2 Result

Below the  $\mathbb{L}_p$ -norm is associated to the random process generated by the SMC algorithm. We also use the abuse of notation  $G_{n,\epsilon}(x) = \phi\left(\frac{x-y_n}{\epsilon}\right)$ ,  $x \in \mathbb{R}^{d_\nu}$ , to represent the incremental weights of the SMC algorithm (that is as described in Section 3.4). Note that, in comparison to (10), we resample at every time-point, so we can use the incremental weights in the estimate, instead of the normalized weights. Note that

$$\Xi_n[\epsilon, N, \mathcal{V}_n, y_{0:n}] = \sum_{i=1}^N \frac{G_{n,\epsilon}(x_n^i)}{\sum_{j=1}^N G_{n,\epsilon}(x_n^j)} V_{n,\epsilon}^N(x_n^i) \quad (13)$$

where  $\Xi_n[\epsilon, N, \mathcal{V}_n, y_{0:n}]$  is the quantity that we discussed in Section 1; from herein we use the R.H.S. of (13) to denote the SMC estimate.



**Theorem 4.1.** *Assume (A1). There exist a  $C < +\infty$  and for any  $p \geq 1$  there exist a  $a_p < \infty$  such that for any  $\epsilon > 0$ ,  $N \geq 1$ ,  $n \geq 1$  and  $y_{0:n}$ :*

$$\left\| \sum_{i=1}^N \frac{G_{n,\epsilon}(x_n^i)}{\sum_{j=1}^N G_{n,\epsilon}(x_n^j)} V_{n,\epsilon}^N(x_n^i) - \mathbb{E}[\mathcal{V}_n(X_{0:n})|y_{0:n}] \right\|_p \leq (n+1) \left[ \frac{a_p \bar{\delta}(\epsilon)}{\sqrt{N}} + C\epsilon \right]$$

where  $\bar{\delta}(\epsilon) = \max\{\delta(\epsilon) \max\{\delta(\epsilon)^2, \delta(\epsilon)^4\}, \delta(\epsilon)^3\}$ , with  $\delta(\epsilon)$  as in (A1) (3) and  $\mathbb{E}[\cdot|y_{0:n}]$  is the expectation w.r.t. the joint smoothing distribution.

*Proof.* After adding and subtracting  $\mathbb{E}^\epsilon[\mathcal{V}_n(X_{0:n})|y_{0:n}]$  one can apply Minkowski followed by Theorem A.1 and Theorem 2.1 to conclude.  $\square$

**Remark 4.1.** *The bound is decomposed into two sources of error. For the SMC approximation, the error tends to decrease as  $\epsilon$  grows as one would expect. Conversely, the ABC error term  $|\mathbb{E}^\epsilon[\mathcal{V}_n(x_{0:n})|y_{0:n}] - \mathbb{E}[\mathcal{V}_n(x_{0:n})|y_{0:n}]|$  grows as  $\epsilon$  grows. Both error rates increase at most linearly with the time parameter. When  $\mathcal{V}_n(x_{0:n}) = \sum_{p=0}^n v_p(x_p)/(n+1)$ , one can remove the linear decay in the bias term. In addition, with  $N$  fixed, the SMC error can be shown to decrease with  $n$ ; see [15, 16].*

## 5 Simulations

### 5.1 Model

Our numerical studies are implemented on the following HMM also considered in, for example, [2]. We take  $d_x = d_y = d$  and the model is:

$$\begin{aligned} X_n &= \frac{X_{n-1}}{2} + \frac{25X_{n-1}}{1+X_{n-1}^2} + 8 \cos(1.2n) + \zeta_{X,n}, \quad n \geq 1, \\ Y_n &= \frac{X_n^2}{20} + \zeta_{Y,n} \quad n \geq 0, \end{aligned}$$

with  $\zeta_{X,n} \stackrel{\text{i.i.d.}}{\sim} \mathcal{N}_d(0, \sigma_X^2 \mathbf{I}_d)$  and independently  $\zeta_{Y,n} \stackrel{\text{i.i.d.}}{\sim} \mathcal{N}_d(0, \sigma_Y^2 \mathbf{I}_d)$  and  $X_0 = 0_d$  with  $0_d$  the  $d$ -dimensional zero vector,  $\mathbf{I}_d$  the  $d \times d$  identity matrix and  $\mathcal{N}_d(\mu, \Sigma)$  the  $d$ -dimensional normal distribution of mean  $\mu$  and covariance matrix  $\Sigma$ . Whilst the conditional density of the observations given the state is not intractable, it will facilitate an investigation into the accuracy of ABC. In this scenario one can obtain an approximation of the ‘correct’ answers using SMC/PMCMC with many particles/iterations.

The objective of our numerical study, for smoothing is to consider the accuracy of ABC, when only considering forward only smoothing (the performance of forward only smoothing relative to using the path of particles has been studied elsewhere - for example [15]). We also want to investigate the worth of RSMC in the ABC context; recall the asymptotic improvements predicted in [14]. Along the way we also consider the issue of the dimension of the HMM and the utility of using ABC in high-dimensions. Finally, the time dependence of the errors are presented, to allow some investigation into Theorem 4.1. When considering PMCMC, we are concerned with both the accuracy of ABC for batch static parameter estimation and the worth of including forward only smoothing as a ‘post-processing’ of the MCMC output.

### 5.2 Smoothing

#### 5.2.1 Implementation Details

We consider estimating the expected mean state over the observation period  $[0, 100]$ ; i.e.  $v_p(x_p) = x_p/101$ ,  $p \in \{0, \dots, 100\}$ . We set  $\sigma_X^2 = 10$  and  $\sigma_Y^2 = 1$ . The data are simulated from the true model with the given parameter values. To obtain a true answer with which to understand the accuracy of the methods we investigate, we use the mean estimate obtained over 50 implementations of the forward smoothing procedure, targeting the exact model, with 5000 particles.

The algorithms for SMC (that is, approximating the exact model) and SMC ABC are run for 10 different values of  $N \in \{100, 200, \dots, 1000\}$  which are labelled  $N_1, \dots, N_{10}$  in the Figures. The SMC ABC approach, i.e. that dynamically resamples, does so when the effective sample size drops below  $N/2$ . For the SMC, SMC ABC and RSMC (which targets the ABC approximation) the hidden state dynamics are used as proposals. We also run the

algorithms for  $d \in \{1, 2, 5, 10\}$  which will also allow us to assess the accuracy of SMC for an ABC HMM in ‘high’ dimensions. To investigate the accuracy of ABC, we compute a true value for  $\mathbb{E}[\mathcal{V}_n(X_{0:n})|y_{0:n}]$ , as discussed above, and then average the  $\mathbb{L}_1$ -error of the estimate  $\Xi_n[\epsilon, N, \mathcal{V}_n, y_{0:n}]$ , calculated with respect to the computed true value, across its dimension, i.e.  $e_n^{N,\epsilon} = \frac{1}{d} |\mathbb{E}[\mathcal{V}_n(X_{0:n})|y_{0:n}] - \Xi_n[\epsilon, N, \mathcal{V}_n, y_{0:n}]|$ , with  $|\cdot|$  the  $\mathbb{L}_1$ -distance. An SMC procedure targeting the exact HMM is also run, to provide some benchmark performance; the corresponding error,  $e_n^N$ , is similarly calculated as the dimension-averaged  $\mathbb{L}_1$ -error of the SMC estimate  $\Xi_n[N, \mathcal{V}_n, y_{0:n}]$  with respect to the same true value as above. All results are averaged over 50 independent runs.

For the ABC specification, we set  $\phi\left(\frac{u-y}{\epsilon}\right) = \mathbb{I}_{\{|u-y|/\epsilon < 1\}}(u)$ ; this will allow us to easily understand the impact of the RSMC. In the implementations of the two SMC ABC schemes described in Section 3.4  $\epsilon$  is set to be the smallest obtainable in a preliminary set of runs. That is, the smallest  $\epsilon$  for which the weights do not become zero at any time-point.

To conclude the numerical study, we consider the time-dependence of the bias. We consider the SMC and ABC using only forward only smoothing as the time parameter increases from 10, 20,  $\dots$ , 100,  $d \in \{1, 2, 5, 10\}$ . The SMC algorithm is run with  $N = 1000$ . RSMC is not considered.

## 5.2.2 Results

The exact and ABC forward smoothing errors,  $e_n^N$  and  $e_n^{N,\epsilon}$  respectively, are presented in Figure 1; the mean errors obtained across the 50 runs are displayed, along with their standard errors. Under the ABC HMM, as one would expect, in almost all of the plots the accuracy of the estimate  $e_n^{N,\epsilon}$  cannot improve with increasing  $N$  (as the bias persists), but the variability of the estimates falls - i.e. the SMC component of  $e_n^{N,\epsilon}$  is being controlled. In the plots for  $d \in \{1, 2\}$ , the exact implementation outperforms, as one would expect, the ABC approximation in terms of accuracy. This is illustrated by the means and standard errors of the smoothing errors  $e_n^N$  being smaller than those of the ABC smoothing errors  $e_n^{N,\epsilon}$  for a vast majority of the values of  $N$ . Interestingly, as the dimension increases ( $d \in \{5, 10\}$ ), the ABC estimates appear to be *more* accurate than their SMC counterparts (at least for this function). One might explain this as follows. For SMC in high-dimensions, one often requires  $N = \mathcal{O}(\kappa^d)$  ( $\kappa > 1$ ) for some stability, but this is not the case for ABC - see [5] and the references therein. These (empirical) results suggest that ABC is a viable approximation technique in higher-dimensions, where it can be difficult to find SMC techniques that always work well.

In Figures 2 and 3 we compare the performance of SMC and RSMC for performing estimation of the smoothing expectation under the ABC HMM. Figure 2 presents the mean and standard errors of the ABC smoothing errors that correspond to estimates calculated using the SMC method and the RSMC method; distinction is made through a further subscript, with the ABC smoothing errors corresponding to the RSMC estimates being denoted  $e_n^{N,\epsilon,R}$ . For clarity of presentation, we only display the results of 7 of the 10 values of  $N$  which were run. From Figure 2, it is noted that the accuracy of the SMC and RSMC procedures for performing ABC forward smoothing are very comparable, with the RSMC estimates even appearing to offer a marginal improvement over the SMC estimates in terms of mean smoothing error. The standard errors of  $e_n^{N,\epsilon}$  and  $e_n^{N,\epsilon,R}$  are more clearly presented in Figure 3. This figure shows that, under the ABC HMM, the variability of the RSMC procedure seems to be slightly less than that of the SMC procedure, especially as  $N$  is allowed to grow. In addition, the observed run times for the ABC forward smoothing procedure were consistently lower when using the RSMC method against the SMC ABC approach. These results suggest, at least under the criteria considered, that the use of RSMC would not only be a viable alternative, but it could be preferable to using SMC with dynamic resampling. This is when using forward smoothing to perform inference with respect to the ABC approximation of the HMM.

In Figure 4 we consider the time dependence of the error  $e_n^{N,\epsilon}$  associated with the SMC method applied to the ABC HMM. We can observe that in this scenario, there is not any obvious increase in the overall error  $e_n^{N,\epsilon}$ , with time, for this particular estimate associated to the smoothing distribution. This is consistent with our theoretical results which illustrate that the error does not grow any worse than linearly with time. As expected, on the basis of the results above, the quality of the SMC approximation appears to deteriorate (for  $N$  fixed) as the dimension grows, but such a deterioration is less obvious for the ABC approximation.

## 5.3 PMMH

### 5.3.1 Implementation Details

As for the smoothing, we estimate the expected mean state over the observation period  $[0, 100]$  as well as estimating the static parameters  $\theta = (\sigma_X, \sigma_Y)$ , with priors as in [2]. We set  $d = 1$  throughout and the data are the same as for the smoothing experiment (when  $d = 1$ ). To obtain a proxy for the true value we ran a PMMH algorithm as

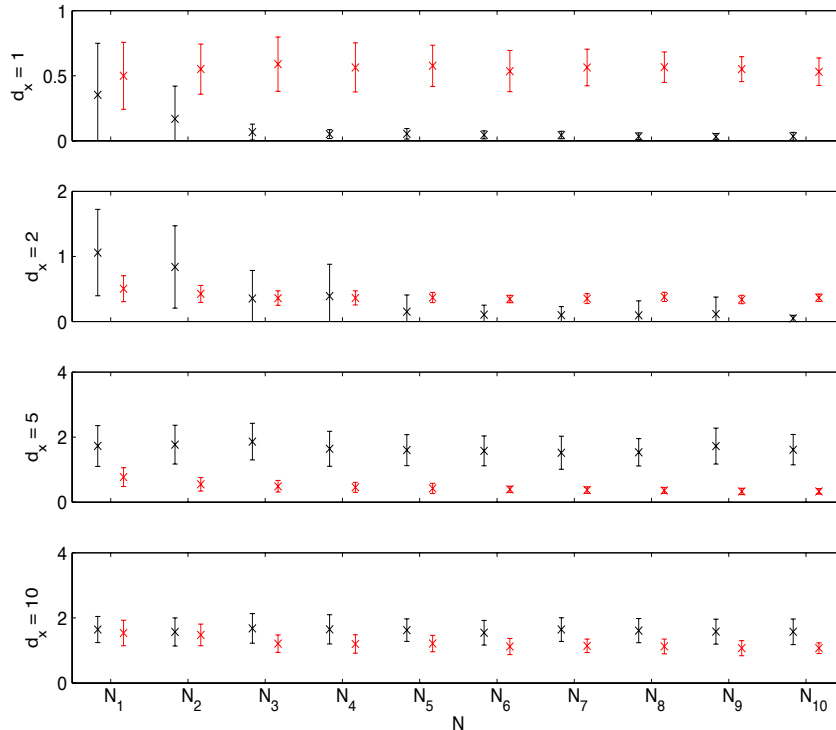


Figure 1: Mean and standard deviations of the smoothing errors associated with the estimates of the mean state ( $v_p(x_p) = x_p/101$ ) obtained over 50 independent implementations of the forward smoothing procedure targeting the true HMM ( $e_n^N$ , black) and its ABC approximation ( $e_n^{N,\epsilon}$ , red). The horizontal axis represents the 10 different values of  $N$  for which we ran both algorithms.

described in [2] for 50000 iterations with 20000 particles (no forward smoothing) and averaged the results over 50 runs. When no forward smoothing is used, only the selected particle (see Section 3.5) is used for the estimate of a smoothed additive functional.

The ABC approximation was as for the smoothing example (that is, the function  $\phi\left(\frac{u-y}{\epsilon}\right) = \mathbb{I}_{\{u:|u-y/\epsilon|<1\}}(u)$ ). To allow direct comparison to running an exact PMMH algorithm (that is, one which uses a dynamic resampling SMC algorithm on the true HMM) we only adopt an SMC ABC algorithm, i.e. we do not consider the use of RSMC here. For the SMC and SMC ABC the hidden state dynamics are used as proposals. The PMMH proposal on the parameters is as in [2]. We run the algorithms for 50000 iterations with a 10000 iteration burn in. In addition, 5 different values of  $N$  are considered  $N \in \{100, 200, \dots, 500\}$  for the forward only smoothing approaches. In comparing to PMMH algorithms that do not use all the particles (and hence the computational cost of the SMC algorithm is  $\mathcal{O}(N)$ ) a number of particles with similar computational costs are run; these were  $\{4427, 17139, 39020, 68258, 107007\}$ .

As with SMC smoothing, the accuracy of the PMMH procedures in estimating the smoothing expectation  $\mathbb{E}[\mathcal{V}_n(X_{0:n})|y_{0:n}]$  is measured using  $e_n^N$  and  $e_n^{N,\epsilon}$ . As above, these errors are calculated as the (dimension-averaged)  $\mathbb{L}_1$ -errors of the PMMH estimates under the exact and ABC HMM, respectively. All results are repeated over 50 independent runs.

### 5.3.2 Results

Our results are displayed in Figures 5-7. In Figure 5, we can observe the accuracy of PMMH estimation of the smoothed additive functional, using SMC updates both with and without forward smoothing, under both the exact HMM and its ABC approximation. Here we observe the expected pattern; the use of forward only smoothing in the PMMH update scheme significantly enhances estimative accuracy for roughly the same computational cost - the accuracy is better and the variance lower. When using forward smoothing in the SMC update mechanism, we further observe that the ABC HMM can be targeted with reasonable accuracy. Consider the effect of increasing  $N$  on the errors in Figure 5. Interestingly, the improvement in estimation is more evident when using forward only smoothing, even though one expects a PMMH algorithm with more particles to mix better (see e.g. [1]) and thus

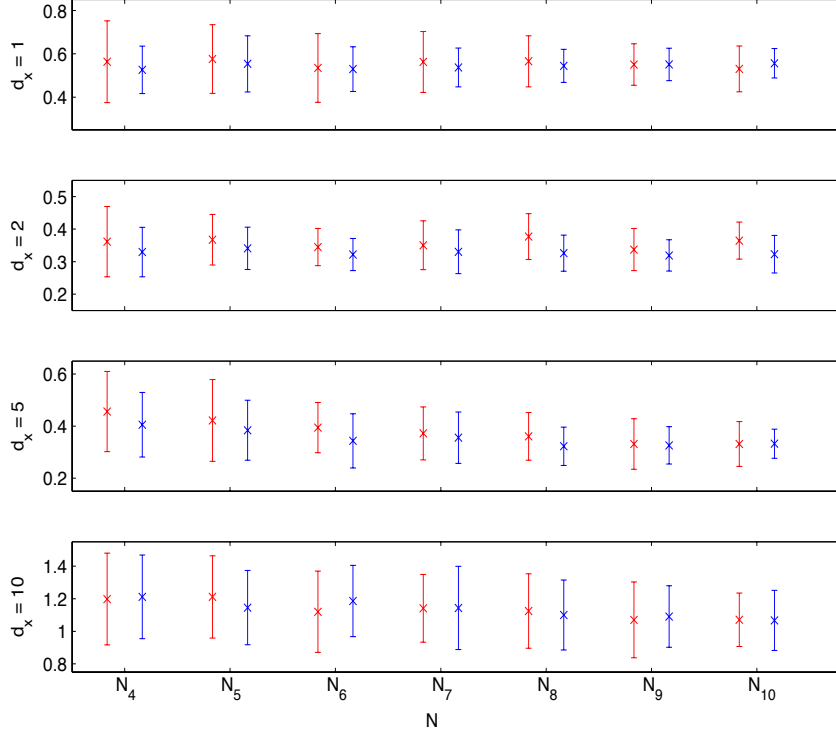


Figure 2: Mean and standard errors of the smoothing errors associated with the estimates of the mean state ( $v_p(x_p) = x_p/101$ ) obtained over 50 independent implementations of the forward smoothing SMC ( $e_n^{N,\epsilon}$ , red) and RSMC ( $e_n^{N,\epsilon,R}$ , blue) procedures targeting the ABC approximation of the smoothing distribution.

the estimation to be most likely improved.

In terms of the estimation of parameters, we consider Figures 6 and 7. Here, we are mainly concerned with the quality of parameter estimation under the ABC HMM without forward smoothing - the forward smoothing cannot contribute anything to parameter estimation here. The accuracy of the ABC is, in general quite biased by up-to 40% of the parameter values. The variance is also quite substantial relative to the exact approach.

## 5.4 Conclusions

On the basis of our numerical study we can tentatively conclude the following. For smoothing:

- In higher dimensions, ABC, in terms of accuracy, is competitive with using standard SMC (even if the model is analytically tractable);
- For ABC with  $\phi(\frac{u-y}{\epsilon})$  specified as an indicator function, one would prefer to use the RSMC procedure over an SMC procedure with dynamic resampling.

For the use of PMMH for performing batch parameter estimation, it would appear that, for moderate length time series, using forward only smoothing is not necessarily useful. If one is interested in the estimation of smoothed additive functionals, however, the use of forward smoothing can provide significant improvements (for the same computational cost) in estimative accuracy when compared to the PMMH procedure in [2]. The ABC procedure produces parameter estimates which are perhaps more biased than estimation of smoothed additive functionals, but this is also linked to the fact that the estimation method used is focussed on the latter quantities. These conclusions, of course, cannot be comprehensive as they are model and quantity (w.r.t. estimation) dependent. However, we have seen similar trends in different examples or different parameter settings for the same model.

## 6 Summary

In this article we have investigated smoothing and static parameter estimation for HMMs with intractable likelihoods. We have constructed SMC and PMCMC based-solutions for ABC approximations and investigated the

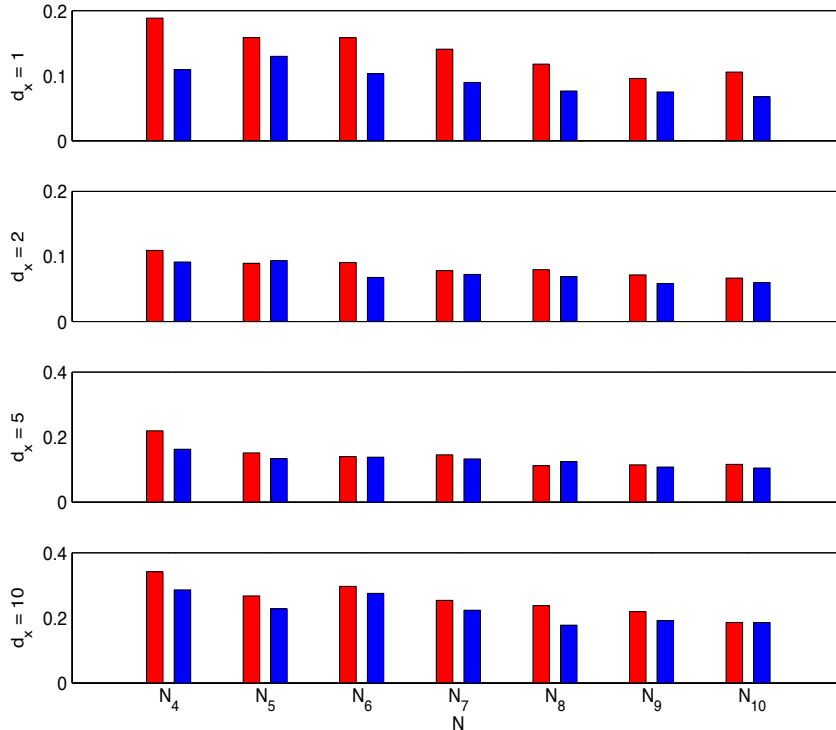


Figure 3: Standard errors of the smoothing errors associated with the estimates of the mean state ( $v_p(x_p) = x_p/101$ ) obtained over 50 independent implementations of the forward smoothing SMC ( $e_n^{N,\epsilon}$ , red) and RSMC ( $e_n^{N,\epsilon,R}$ , blue) procedures targeting the ABC approximation of the smoothing distribution. These standard errors are also displayed in Figure 2.

bias associated to our procedure. There are several extensions to the work that has been considered here. From the perspective of parameter estimation, we have only considered batch estimation by using PMCMC. In many practical problems, one is often interested in performing statistical inference as samples arrive online. We are currently investigating methodology for this problem in [19, 28] and the theoretical and empirical work here is of great relevance in these latter ideas; in particular when applying the online EM algorithm as considered in [28]. In this article we have focussed upon the forward-only smoothing technique in [16], however, this is not the only possibility; one can also investigate the ideas in [6] in the context of ABC. In particular, the relative performance of these procedures is of interest.

### Acknowledgements

The first author was supported by an EPSRC grant. The second author was supported by an MOE grant. The second and fourth authors acknowledge assistance from the LMS research in pairs grant. The second and third authors acknowledge assistance from an EPSRC platform grant EP/I019111/1.

## A SMC Error

We give an analysis of the self-normalized estimate that was not the objective in [16] and is explicit in  $\epsilon$ . To that end, we introduce the following notations, to keep a consistent notation with [15, 16] on which our analysis relies. We set:

$$\begin{aligned}
 H((x, u), (x, u')) &= g(x', u')f(x, x') \\
 G_{n,\epsilon}(u) &= \phi\left(\frac{u - y_n}{\epsilon}\right).
 \end{aligned}$$

To avoid notational overload, we will simply write  $H(x, x')$ ,  $x \in \mathbb{R}^{d_x} \times \mathbb{R}^{d_y} =: E$  and  $G_{n,\epsilon}(x)$ ,  $x \in \mathbb{R}^{d_x} \times \mathbb{R}^{d_y}$  (despite the independence of  $G_{n,\epsilon}$  on only  $\mathbb{R}^{d_x}$ ).

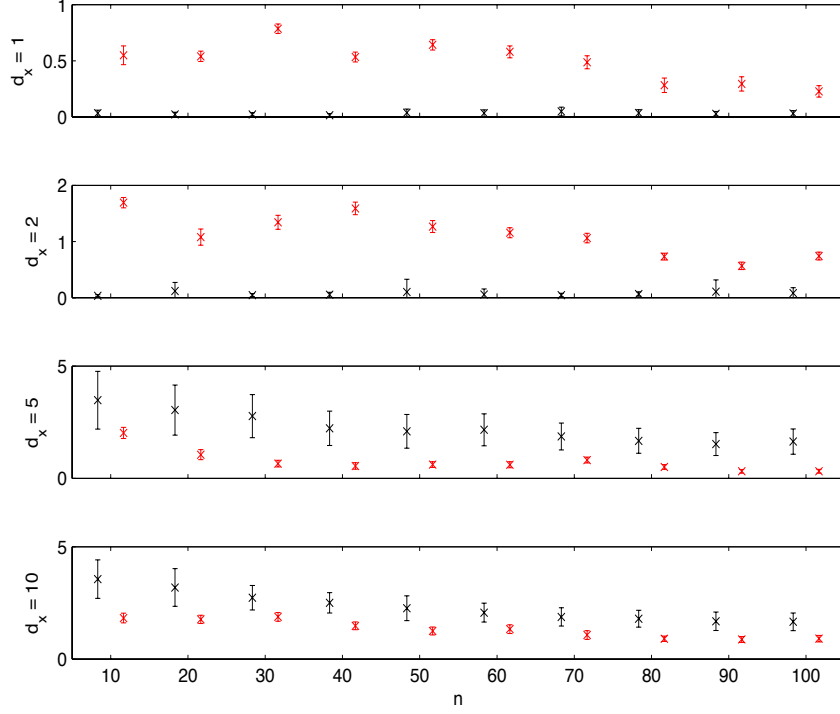


Figure 4: Time dependence (horizontal axis) of the smoothing errors associated with the estimates of the mean state ( $v_p(x_p) = x_p/(n+1)$ ,  $n = 10, \dots, 100$ ) obtained over 50 independent implementations of the forward smoothing SMC ( $e_n^N$ , black) and ABC ( $e_n^{N,\epsilon}$ , red) procedures targeting the true and ABC approximation of the smoothing distribution respectively.

We consider the approximation of the path measure  $\mathbb{Q}_{n,\epsilon}(\mathcal{V}_n)$ , which is by definition:

$$\mathbb{Q}_{n,\epsilon}(\mathcal{V}_n) := \frac{1}{\gamma_{n,\epsilon}(1)} \int_{E^{n+1}} \left[ \prod_{p=0}^{n-1} G_{p,\epsilon}(x_p) \right] \mathcal{V}_n(x_{0:n}) \eta_0(x_0) \prod_{p=1}^n H(x_{p-1}, x_p) dx_{0:p}$$

where

$$\gamma_{n,\epsilon}(f) = \int_{E^{n+1}} \left[ \prod_{p=0}^{n-1} G_{p,\epsilon}(x_p) \right] f(x_n) \eta_0(x_0) \prod_{p=1}^n H(x_{p-1}, x_p) dx_{0:p}$$

and  $f : E \rightarrow \mathbb{R}$ . Recall for additive functionals  $\mathcal{V}_n(x_{0:n}) = \sum_{p=0}^n v_p(x_p)$

$$V_{n,\epsilon}^N(x) = v_n(x) + \sum_{i=1}^N \frac{G_{n-1,\epsilon}(x_{n-1}^i) H_n(x_{n-1}^i, x)}{\sum_{j=1}^N G_{n-1,\epsilon}(x_{n-1}^j) H_n(x_{n-1}^j, x)} V_{n-1,\epsilon}^N(x_{n-1}^i)$$

with  $V_{0,\epsilon}^N = v_0$ ; c.f. (9). We remind the reader that  $v_p$  is a function on  $\mathbb{R}^{d_x}$  only.

**Theorem A.1.** *Assume (A1). Then for any  $1 \leq p < +\infty$ , there exist a  $a_p < +\infty$  such that for any  $\epsilon > 0$ ,  $N \geq 1$ ,  $n \geq 1$ ,  $y_{0:n}$ :*

$$\left\| \sum_{i=1}^N \frac{G_{n,\epsilon}(x_n^i)}{\sum_{j=1}^N G_{n,\epsilon}(x_n^j)} V_{n,\epsilon}^N(x_n^i) - \mathbb{E}^\epsilon[\mathcal{V}_n(X_{0:n}) | y_{0:n}] \right\|_p \leq \frac{a_p(n+1)\bar{\delta}(\epsilon)}{\sqrt{N}}$$

where  $\bar{\delta}(\epsilon) = \max\{\delta(\epsilon) \max\{\delta(\epsilon)^2, \delta(\epsilon)^4\}, \delta(\epsilon)^3\}$ , with  $\delta(\epsilon)$  as in (A1) (3) and  $\mathbb{E}^\epsilon[\cdot | y_{0:n}]$  is the expectation w.r.t. the joint ABC smoothing distribution.

*Proof.* Consider the decomposition:

$$\sum_{i=1}^N \frac{G_{n,\epsilon}(x_n^i)}{\sum_{j=1}^N G_{n,\epsilon}(x_n^j)} V_{n,\epsilon}^N(x_n^i) - \frac{\mathbb{Q}_{n,\epsilon}(G_{n,\epsilon} \mathcal{V}_n)}{\eta_{n,\epsilon}(G_{n,\epsilon})} =$$

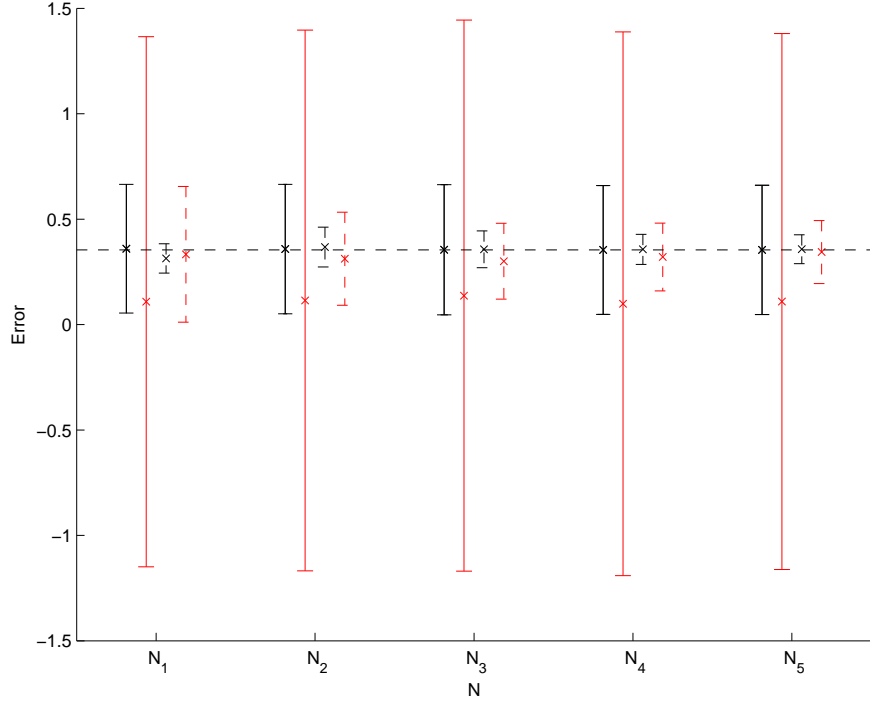


Figure 5: Standard errors of the smoothing errors associated with the estimates of the mean state obtained over 50 independent implementations. The PMCMC with exact SMC is in black, the ABC in red. The dotted lines indicate the usage of forward only smoothing. The dotted horizontal line is the estimated true value.

$$\begin{aligned}
& \frac{\frac{1}{N} \sum_{i=1}^N G_{n,\epsilon}(x_n^i) V_{n,\epsilon}^N(x_n^i)}{\eta_{n,\epsilon}(G_{n,\epsilon}) \frac{1}{N} \sum_{j=1}^N G_{n,\epsilon}(x_n^j)} \left( \eta_{n,\epsilon}(G_{n,\epsilon}) - \frac{1}{N} \sum_{j=1}^N G_{n,\epsilon}(x_n^j) \right) \\
& + \frac{1}{\eta_{n,\epsilon}(G_{n,\epsilon})} \left( \frac{1}{N} \sum_{i=1}^N G_{n,\epsilon}(x_n^i) V_{n,\epsilon}^N(x_n^i) - \mathbb{Q}_{n,\epsilon}(G_{n,\epsilon} \mathcal{V}_n) \right) \tag{14}
\end{aligned}$$

where we recall the normalized  $n$ -time marginal  $\eta_{n,\epsilon}(f) = \gamma_{n,\epsilon}(f)/\gamma_{n,\epsilon}(1)$ . Note that

$$\begin{aligned}
\frac{\mathbb{Q}_{n,\epsilon}(G_{n,\epsilon} \mathcal{V}_n)}{\eta_{n,\epsilon}(G_{n,\epsilon})} &= \frac{1}{\gamma_{n,\epsilon}(G_{n,\epsilon})} \int_{E^{n+1}} \left[ \prod_{p=0}^n G_{p,\epsilon}(x_p) \right] \mathcal{V}_n(x_{0:n}) \eta_0(x_0) \prod_{p=1}^n H(x_{p-1}, x_p) dx_{0:p} \\
&= \mathbb{E}^\epsilon [\mathcal{V}_n(X_{0:n}) | y_{0:n}]
\end{aligned}$$

which is the quantity of interest.

To consider an  $\mathbb{L}_p$ -analysis, we can split the two terms in (14) via Minkowski. We consider the first term:

$$\left\| \frac{\frac{1}{N} \sum_{i=1}^N G_{n,\epsilon}(x_n^i) V_{n,\epsilon}^N(x_n^i)}{\eta_{n,\epsilon}(G_{n,\epsilon}) \frac{1}{N} \sum_{j=1}^N G_{n,\epsilon}(x_n^j)} \left( \eta_{n,\epsilon}(G_{n,\epsilon}) - \frac{1}{N} \sum_{j=1}^N G_{n,\epsilon}(x_n^j) \right) \right\|_p$$

Now, one has:  $V_{n,\epsilon}^N(x_n^i) \leq \sum_{p=0}^n \|v_p\|$ . This is proved by induction, the initialization with  $n = 0$  being obvious, assuming for  $n - 1$ , one has:

$$V_{n,\epsilon}^N(x) = v_n(x) + \sum_{i=1}^N \frac{G_{n-1,\epsilon}(x_{n-1}^i) H_n(x_{n-1}^i, x)}{\sum_{j=1}^N G_{n-1,\epsilon}(x_{n-1}^j) H_n(x_{n-1}^j, x)} V_{n-1,\epsilon}^N(x_{n-1}^i) \leq v_n(x) + \sum_{p=0}^{n-1} \|v_p\|$$

and one easily concludes. Thus, it follows:

$$\left\| \frac{\frac{1}{N} \sum_{i=1}^N G_{n,\epsilon}(x_n^i) V_{n,\epsilon}^N(x_n^i)}{\eta_{n,\epsilon}(G_{n,\epsilon}) \frac{1}{N} \sum_{j=1}^N G_{n,\epsilon}(x_n^j)} \left( \eta_{n,\epsilon}(G_{n,\epsilon}) - \sum_{j=1}^N G_{n,\epsilon}(x_n^j) \right) \right\|_p$$

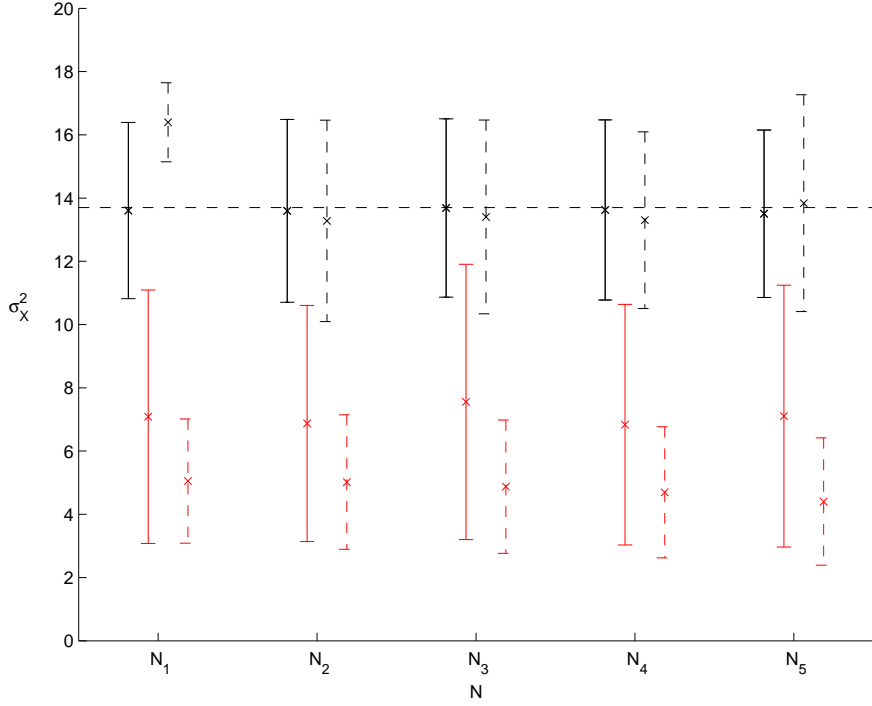


Figure 6: Standard errors of the smoothing errors associated with the estimates of  $\sigma_X^2$  obtained over 50 independent implementations. The PMCMC with exact SMC is in black, the ABC in red. The dotted lines indicate the usage of forward only smoothing. The dotted horizontal line is the estimated true value.

$$\leq (n+1)\bar{v}\underline{\alpha}(\epsilon)^{-1} \left\| \eta_{n,\epsilon}(G_{n,\epsilon}) - \frac{1}{N} \sum_{j=1}^N G_{n,\epsilon}(x_n^j) \right\|_p.$$

By Theorem 7.4.4 of [14] that there exist some  $a_p < \infty$  such that

$$\left\| \left( \eta_{n,\epsilon}(G_{n,\epsilon}) - \frac{1}{N} \sum_{j=1}^N G_{n,\epsilon}(x_n^j) \right) \right\|_p \leq \frac{a_p \delta(\epsilon)^2 \bar{\alpha}(\epsilon)}{\sqrt{N}}$$

hence

$$\left\| \frac{\frac{1}{N} \sum_{i=1}^N G_{n,\epsilon}(x_n^i) V_{n,\epsilon}^N(x_n^i)}{\eta_{n,\epsilon}(G_{n,\epsilon}) \frac{1}{N} \sum_{j=1}^N G_{n,\epsilon}(x_n^j)} \left( \eta_{n,\epsilon}(G_{n,\epsilon}) - \frac{1}{N} \sum_{j=1}^N G_{n,\epsilon}(x_n^j) \right) \right\|_p \leq \frac{a_p(n+1)\delta(\epsilon)^3}{\sqrt{N}}$$

for some  $a_p < \infty$  that does not depend upon  $n$  or  $\epsilon$ . To deal with the second term in (14) one can use Lemma A.1 along with (A1) (3) to conclude.  $\square$

## A.1 Technical Result

We provide a proof of Lemma A.1. To that end, introduce the operator:

$$D_{p,n,\epsilon}^N(V_n)(x_p) = \int \mathcal{M}_{p,\epsilon}^N(x_p, d_{0:p-1}) \mathcal{Q}_{p,n,\epsilon}(x_p, dx_{p:n}) V_n(x_n)$$



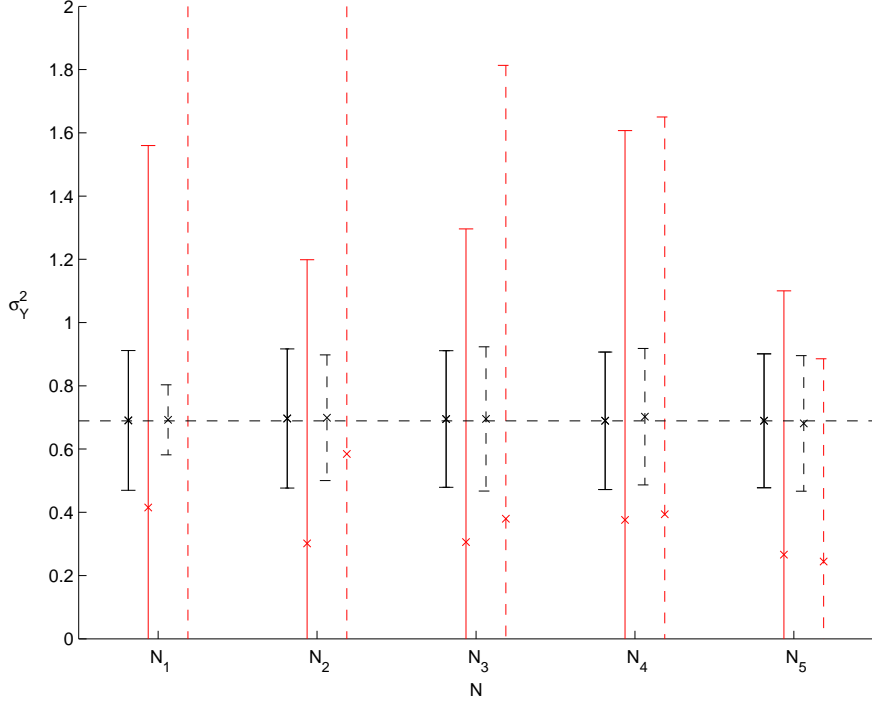


Figure 7: Standard errors of the smoothing errors associated with the estimates of  $\sigma_Y^2$  obtained over 50 independent implementations. The PMCMC with exact SMC is in black, the ABC in red. The dotted lines indicate the usage of forward only smoothing. The dotted horizontal line is the estimated true value.

where

$$\begin{aligned}
\mathcal{M}_{p,\epsilon}^N(x_p, dx_{0:p-1}) &= \prod_{q=1}^p M_{q,\eta_{q-1}^N,\epsilon}(x_q, dx_{q-1}) \\
\mathcal{Q}_{p,n,\epsilon}(x_p, dx_{p:n}) &= \prod_{q=p+1}^n Q_{q,\epsilon}(x_{q-1}, dx_q) \\
M_{q,\eta_{q-1}^N,\epsilon}(x_q, dx_{q-1}) &= \frac{\eta_{q-1}^N(dx_{q-1})G_{q-1,\epsilon}(x_{q-1})H_q(x_{q-1}, x_q)}{\eta_{q-1}^N(G_{q-1,\epsilon}H_q(\cdot, x_q))} \\
Q_{q,\epsilon}(x_{q-1}, dx_q) &= G_{q-1,\epsilon}(x_{q-1})H_q(x_{q-1}, x_q)dx_q \\
\eta_{q-1}^N(dx_{q-1}) &= \frac{1}{N} \sum_{i=1}^N \delta_{x_{q-1}^i}(dx_{q-1})
\end{aligned}$$

where all the conventions of [15] are preserved. In the empirical measure in the final line, one considers the mutated particles. We also use the convention  $Q_{p,n,\epsilon} = Q_{p+1,\epsilon} \dots Q_{n,\epsilon}$ . Note for the backward kernel, when considering the filter  $\hat{\eta}_{q-1,\epsilon}$  one can write

$$M_{q,\hat{\eta}_{q-1,\epsilon},\epsilon}(x_q, dx_{q-1}) = \frac{\hat{\eta}_{q-1,\epsilon}(dx_{q-1})H_q(x_{q-1}, x_q)}{\hat{\eta}_{q-1,\epsilon}(H_q(\cdot, x_q))}.$$

**Lemma A.1.** *Assume (A1). Then for any  $1 \leq p < +\infty$  there exist a  $a_p < +\infty$  such that for any  $\epsilon > 0$ ,  $N \geq 1$ ,  $n \geq 1$ ,  $y_{0:n}$ :*

$$\left\| \frac{1}{N} \sum_{i=1}^N G_{n,\epsilon}(x_n^i) V_{n,\epsilon}^N(x_n^i) - \mathbb{Q}_{n,\epsilon}(G_{n,\epsilon} \mathcal{V}_n) \right\|_p \leq \frac{a_p(n+1)\tilde{\delta}(\epsilon)\bar{\alpha}(\epsilon)}{\sqrt{N}}$$

where  $\tilde{\delta}(\epsilon) = \max\{\delta(\epsilon)^2, \delta(\epsilon)^4\}$ , with  $\delta(\epsilon), \bar{\alpha}(\epsilon)$  as in (A1) (3).

*Proof.* The proof of this result follows the proof of Theorem 3.2 of [15]. The complications are the control of the oscillations of  $P_{p,n,\epsilon}^N(G_{n,\epsilon}\mathcal{V}_n) = D_{p,n,\epsilon}^N(G_{n,\epsilon}\mathcal{V}_n)/D_{p,n,\epsilon}^N(1)$  with a different function as well as obtaining a rate w.r.t.  $\epsilon$ . It is remarked that the application of the Kintchine-type inequality in Theorem 3.2 of [15] will not add any dependence upon  $\epsilon$ .

Using the definition of  $D_{p,n,\epsilon}^N$  one has

$$D_{p,n,\epsilon}^N(G_{n,\epsilon}\mathcal{V}_n) = Q_{p,n,\epsilon}(G_{n,\epsilon}) \sum_{q=0}^p M_{p,\eta_{p-1}^N,\epsilon} \cdots M_{q+1,\eta_q^N,\epsilon}(v_q) + \sum_{q=p+1}^n Q_{p,q,\epsilon}(v_q Q_{q,n,\epsilon}(G_{n,\epsilon})).$$

Thus it follows that:

$$P_{p,n,\epsilon}^N(G_{n,\epsilon}\mathcal{V}_n) = \frac{Q_{p,n,\epsilon}(G_{n,\epsilon})}{Q_{p,n,\epsilon}(1)} \sum_{q=0}^{p-1} M_{p,\eta_{p-1}^N,\epsilon} \cdots M_{q+1,\eta_q^N,\epsilon}(v_q) + \sum_{q=p}^n \frac{S_{p,q,\epsilon}(\overline{Q}_{q,n,\epsilon}(G_{n,\epsilon})v_q)}{S_{p,q,\epsilon}(\overline{Q}_{q,n,\epsilon}(1))} \quad (15)$$

where  $S_{p,q,\epsilon}(v) = Q_{p,q,\epsilon}(v)/Q_{p,q,\epsilon}(1)$  and  $\overline{Q}_{q,n,\epsilon}(v) = Q_{q,n,\epsilon}(v)/\eta_{q,\epsilon}(Q_{q,n,\epsilon}(1))$ . To deal with oscillations of  $P_{p,n,\epsilon}^N(G_{n,\epsilon}\mathcal{V}_n)$  we consider both sums separately.

We begin with the first term on the R.H.S. of (15). In particular the difference with arguments  $x$  and  $y$ :

$$\frac{Q_{p,n,\epsilon}(G_{n,\epsilon})(x)}{Q_{p,n,\epsilon}(1)(x)} \sum_{q=0}^{p-1} M_{p,\eta_{p-1}^N,\epsilon} \cdots M_{q+1,\eta_q^N,\epsilon}(v_q)(x) - \frac{Q_{p,n,\epsilon}(G_{n,\epsilon})(y)}{Q_{p,n,\epsilon}(1)(y)} \sum_{q=0}^{p-1} M_{p,\eta_{p-1}^N,\epsilon} \cdots M_{q+1,\eta_q^N,\epsilon}(v_q)(y).$$

Then, treating the summands, one has

$$\begin{aligned} & \frac{Q_{p,n,\epsilon}(G_{n,\epsilon})(x)}{Q_{p,n,\epsilon}(1)(x)} \left[ M_{p,\eta_{p-1}^N,\epsilon} \cdots M_{q+1,\eta_q^N,\epsilon}(v_q)(x) - M_{q+1,\eta_q^N,\epsilon}(v_q)(y) \right] + \\ & M_{q+1,\eta_q^N,\epsilon}(v_q)(y) \left[ S_{p,n,\epsilon}(G_{n,\epsilon})(x) - S_{p,n,\epsilon}(G_{n,\epsilon})(y) \right] \end{aligned}$$

which is clearly upper-bounded by

$$2\|G_{n,\epsilon}\| \|v_q\| [\beta(M_{p,\eta_{p-1}^N,\epsilon} \cdots M_{q+1,\eta_q^N,\epsilon}) + \beta(S_{p,n,\epsilon})]. \quad (16)$$

Now consider the second term on the R.H.S. of (15). In particular, for each summand, one has after subtracting the function in  $y$  from that in  $x$

$$\begin{aligned} & \frac{S_{p,q,\epsilon}(\overline{Q}_{q,n,\epsilon}(G_{n,\epsilon})v_q)(x) - S_{p,q,\epsilon}(\overline{Q}_{q,n,\epsilon}(G_{n,\epsilon})v_q)(y)}{S_{p,q,\epsilon}(\overline{Q}_{q,n,\epsilon}(1)(x))} \\ & + S_{p,q,\epsilon}(\overline{Q}_{q,n,\epsilon}(G_{n,\epsilon})v_q)(y) \left[ \frac{S_{p,q,\epsilon}(\overline{Q}_{q,n,\epsilon}(1)(y)) - S_{p,q,\epsilon}(\overline{Q}_{q,n,\epsilon}(1)(x))}{S_{p,q,\epsilon}(\overline{Q}_{q,n,\epsilon}(1)(y))S_{p,q,\epsilon}(\overline{Q}_{q,n,\epsilon}(1)(x))} \right] \end{aligned}$$

Dealing with the two terms separately, one can easily show that the first term is upper-bounded by  $2b_{q,n,\epsilon}^2 \|v_q\| \|G_{n,\epsilon}\| \beta(S_{p,q,\epsilon})$  where  $b_{q,n,\epsilon} = \sup_{x,y} Q_{p,n,\epsilon}(1)(x)/Q_{p,n,\epsilon}(1)(y)$ . Similarly using trivial manipulations, one can also show that the second term is upper-bounded by same term, yielding the upper-bound

$$4b_{q,n,\epsilon}^2 \|v_q\| \|G_{n,\epsilon}\| \beta(S_{p,q,\epsilon}) \quad (17)$$

Combining the bounds (16)-(17) one can deduce that:

$$\begin{aligned} \text{Osc}(P_{p,n,\epsilon}^N(G_{n,\epsilon}\mathcal{V}_n)) & \leq 2\|G_{n,\epsilon}\| \sum_{q=0}^{p-1} \|v_q\| [\beta(M_{p,\eta_{p-1}^N,\epsilon} \cdots M_{q+1,\eta_q^N,\epsilon}) + \beta(S_{p,n,\epsilon})] \\ & + 4\|G_{n,\epsilon}\| \sum_{q=0}^{p-1} \|v_q\| b_{q,n,\epsilon}^2 \beta(S_{p,q,\epsilon}). \end{aligned}$$

Thus, one has, via the proof of Theorem 3.2 of [15]:

$$\left\| \frac{1}{N} \sum_{i=1}^N G_{n,\epsilon}(x_n^i) V_{n,\epsilon}^N(x_n^i) - \mathbb{Q}_{n,\epsilon}(G_{n,\epsilon}\mathcal{V}_n) \right\|_p \leq a_p \sum_{p=0}^n b_{p,n,\epsilon}^2 c_{p,n,\epsilon}^N$$

where

$$c_{p,n,\epsilon}^N = \mathbb{E} \left[ 2 \|G_{n,\epsilon}\| \sum_{q=0}^{p-1} \|v_q\| [\beta(M_{p,\eta_{p-1}^N,\epsilon} \cdots M_{q+1,\eta_q^N,\epsilon}) + \beta(S_{p,n,\epsilon})] + 4 \|G_{n,\epsilon}\| \sum_{q=0}^{p-1} \|v_q\| b_{q,n,\epsilon}^2 \beta(S_{p,q,\epsilon}) \right].$$

To complete the proof we need to consider the term  $c_{p,n,\epsilon}^N$ ; to that end, we quote the following bounds which follow from (A1); see [15] and the citations therein for details:

$$b_{p,n,\epsilon} \leq \delta(\epsilon)\rho^4 \quad \beta(S_{p,q,\epsilon}) \leq (1 - \delta(\epsilon)^{-1}\rho^{-8})^{q-p} \quad \beta(M_{p,\eta_{p-1}^N,\epsilon} \cdots M_{q+1,\eta_q^N,\epsilon}) \leq (1 - \rho^{-8})^{p-q}.$$

First consider the expression

$$2 \|G_{n,\epsilon}\| \sum_{p=0}^n b_{p,n,\epsilon}^2 \sum_{q=0}^{p-1} \|v_q\| [\beta(M_{p,\eta_{p-1}^N,\epsilon} \cdots M_{q+1,\eta_q^N,\epsilon}) + \beta(S_{p,n,\epsilon})]$$

which is upper-bounded by

$$2\bar{\alpha}(\epsilon)(\delta(\epsilon)\rho^4)^{2\bar{v}} \sum_{p=0}^n \sum_{q=0}^{p-1} [(1 - \rho^{-8})^{p-q} + (1 - \delta(\epsilon)^{-1}\rho^{-8})^{n-p}]$$

with  $\bar{v}$  as in (A1) (4). By standard manipulations, this is upper-bounded by

$$C\bar{\alpha}(\epsilon)\delta(\epsilon)^2(n+1)$$

for  $C < \infty$  that does not depend upon  $n$  or  $\epsilon$ . Second the expression

$$\sum_{p=0}^n b_{p,n,\epsilon}^2 4 \|G_{n,\epsilon}\| \sum_{q=0}^{p-1} \|v_q\| b_{q,n,\epsilon}^2 \beta(S_{p,q,\epsilon})$$

which is upper-bounded by

$$4\bar{\alpha}(\epsilon)(\delta(\epsilon)\rho^4)^{4\bar{v}} \sum_{p=0}^n \sum_{q=p}^n (1 - \delta(\epsilon)^{-1}\rho^{-8})^{q-p}.$$

Again, by standard manipulations one can upper-bound this latter expression by

$$C\bar{\alpha}(\epsilon)\delta(\epsilon)^4(n+1)$$

for  $C < \infty$  that does not depend upon  $n$  or  $\epsilon$ ; we can now conclude.  $\square$

## B ABC Error

Below we will repeatedly apply Theorem 2 of [22]. This can also be established under (A1) for smoothed ABC; the proof is omitted and follows the description in [22]. Recall that the ABC approximation of the joint smoothing density is:

$$\hat{\eta}_{n,\epsilon}(x_{0:n}) = \frac{\prod_{i=0}^n \int_{\mathbb{R}^{d_y}} \phi\left(\frac{u-y_i}{\epsilon}\right) g(x_i, u) du \eta_0(x_0) \prod_{i=1}^n f(x_{i-1}, x_i)}{\int_{\mathbb{R}^{(n+1)d_x}} \prod_{i=0}^n \int_{\mathbb{R}^{d_y}} \phi\left(\frac{u-y_i}{\epsilon}\right) g(x_i, u) du \eta_0(x_0) \prod_{i=1}^n f(x_{i-1}, x_i) dx_{0:n}}.$$

It is stressed that the analysis here is performed by integrating the auxiliary data, whilst the SMC analysis works on the joint space of auxiliary data and hidden state.

*Proof of Theorem 2.1.* We have

$$|\mathbb{E}[\mathcal{V}_n(X_{0:n})|y_{0:n}] - \mathbb{E}^\epsilon[\mathcal{V}_n(X_{0:n})|y_{0:n}]| = \left| \sum_{p=0}^n \int v_p(x_p) [\hat{\eta}_n(x_{p:n}) - \hat{\eta}_{n,\epsilon}(x_{p:n})] dx_{p:n} \right|$$

where  $\hat{\eta}_{p,\epsilon}(x_{p:n})$  and  $\hat{\eta}_p(x_{p:n})$  are the ABC and true smoothers. Using the backward representation of the smoothers, one has the decomposition of the R.H.S.:

$$\left| \sum_{p=0}^n \int v_p(x_p) [\hat{\eta}_n(x_{p:n}) - \hat{\eta}_{n,\epsilon}(x_{p:n})] dx_{p:n} \right| = \left| \sum_{p=0}^n \hat{\eta}_n M_{n:p+1, \hat{\eta}_{n-1:p}}(v_p) - \hat{\eta}_n^\epsilon M_{n:p+1, \hat{\eta}_{n-1:p, \epsilon}}(v_p) \right| \quad (18)$$

where

$$\begin{aligned}\hat{\eta}_n M_{n:p+1, \hat{\eta}_{n-1:p}}(v_p) &= \int \hat{\eta}_n(x_n) \prod_{q=p+1}^n M_{q, \hat{\eta}_{q-1}}(x_q, dx_{q-1}) v_p(x_p) dx_n \\ \hat{\eta}_{n, \epsilon} M_{n:p+1, \hat{\eta}_{n-1:p, \epsilon}}(v_p) &= \int \hat{\eta}_{n, \epsilon}(x_n) \prod_{q=p+1}^n M_{q, \hat{\eta}_{q-1, \epsilon}}(x_q, dx_{q-1}) v_p(x_p) dx_n\end{aligned}$$

with the backward kernels:

$$M_{q, \hat{\eta}_{q-1}}(x_q, dx_{q-1}) = \frac{\hat{\eta}_{q-1}(x_{q-1}) f(x_{q-1}, x_q)}{\hat{\eta}_{q-1}(f(\cdot, x_q))} dx_{q-1} \quad (19)$$

$$M_{q, \hat{\eta}_{q-1, \epsilon}}(x_q, dx_{q-1}) = \frac{\hat{\eta}_{q-1, \epsilon}(x_{q-1}) f(x_{q-1}, x_q)}{\hat{\eta}_{q-1, \epsilon}(f(\cdot, x_q))} dx_{q-1}. \quad (20)$$

We will drop the  $\eta$  subscripts for the remainder of the proof.

One can now adopt a telescoping sum decomposition for each summand of the R.H.S. of (18):

$$\sum_{s=1}^{n-p} \left( \hat{\eta}_n M_{n:n-s, \epsilon} [M_{n-s+1} - M_{n-s+1, \epsilon}] (M_{n-s:p+1}(v_p)) \right) + [\hat{\eta}_n - \hat{\eta}_{n, \epsilon}] (M_{n:p+1, \epsilon}(v_p))$$

For the second term, one can use Theorem 2 of [22]. Thus concentrating on the summands in the first term we have

$$\begin{aligned}\hat{\eta}_n M_{n:n-s, \epsilon} [M_{n-s+1} - M_{n-s+1, \epsilon}] (M_{n-s:p+1}(v_p)) &\leq \|\hat{\eta}_n M_{n:n-s, \epsilon} [M_{n-s+1} - M_{n-s+1, \epsilon}]\|_{TV} \times \\ &\quad \left( \prod_{q=n-s}^{p+1} \beta(M_q) \right) \text{Osc}(v_p)\end{aligned}$$

via (A1) (1), (4) and Lemma B.1 it clearly follows that there exist a  $C < \infty$  and  $\xi \in (0, 1)$  which do not depend upon  $n, y_{0:n}$   $\epsilon$  such that

$$\hat{\eta}_n M_{n:n-s, \epsilon} [M_{n-s+1} - M_{n-s+1, \epsilon}] (M_{n-s:p+1}(v_p)) \leq C \epsilon \xi^{p+s+2-n}.$$

As a result, we have

$$|\mathbb{E}[\mathcal{V}_n(X_{0:n})|y_{0:n}] - \mathbb{E}^\epsilon[\mathcal{V}_n(X_{0:n})|y_{0:n}]| \leq C \sum_{p=0}^n \left[ \left( \sum_{s=1}^{n-p} \xi^{p+s+2-n} \right) + 1 \right] \epsilon$$

for  $C < \infty$  that does not depend upon  $n, y_{0:n}$   $\epsilon$ . Elementary manipulations allow us to conclude.  $\square$

## B.1 Technical Result

**Lemma B.1.** *Assume (A1). Then there exist a  $C < +\infty$  such that for any  $k \in \{0, \dots, n-2\}$   $\epsilon > 0$ ,  $y_{0:n}$  and  $\varphi \in \mathcal{B}_b(\mathbb{R}^{d_x})$  we have*

$$\sup_{x \in \mathbb{R}^{d_x}} \left| \int [M_{k+1, \hat{\eta}_{k, \epsilon, \epsilon}}(x, dz) - M_{k+1, \hat{\eta}_k}(x, dz)] \varphi(z) \right| \leq C \epsilon$$

where  $M_{k+1, \hat{\eta}_{k, \epsilon, \epsilon}}$  and  $M_{k+1, \hat{\eta}_k}$  are defined in (19)-(20) and  $\hat{\eta}_{k, \epsilon}$  and  $\hat{\eta}_k$  are the ABC and true filters.

*Proof.* We have the decomposition:

$$\begin{aligned}\int [M_{k+1, \hat{\eta}_{k, \epsilon, \epsilon}}(x, dz) - M_{k+1, \hat{\eta}_k}(x, dz)] \varphi(z) &= \int \varphi(z) f(z, x) \left[ \frac{\hat{\eta}_{k, \epsilon}(z) - \hat{\eta}_k(z)}{\int \hat{\eta}_{k, \epsilon}(u) f(u, x) du} \right. \\ &\quad \left. + \hat{\eta}_k(z) \left\{ \frac{\int [\hat{\eta}_k(u) - \hat{\eta}_{k, \epsilon}(u)] f(u, x) du}{\int \hat{\eta}_{k, \epsilon}(u) f(u, x) du \int \hat{\eta}_k(u) f(u, x) du} \right\} \right] dz\end{aligned}$$

where we have suppressed the data from the notation.

Dealing with the first part, we have for some  $C$  that does not depend upon  $x, k, \epsilon$  or  $y_{0:n}$

$$\int \varphi(z) f(z, x) \left[ \frac{\hat{\eta}_{k,\epsilon}(z) - \hat{\eta}_k(z)}{\int \hat{\eta}_{k,\epsilon}(u) f(u, x) du} \right] dz \leq C\epsilon \|\varphi\|$$

where we have used (A1) (1) in the denominator and to control  $f$  as well as Theorem 2 of [22]. Now, for the second part

$$\int \varphi(z) f(z, x) \hat{\eta}_k(z) \left\{ \frac{\int [\hat{\eta}_k(u) - \hat{\eta}_{k,\epsilon}(u)] f(u, x) du}{\int \hat{\eta}_{k,\epsilon}(u) f(x, u) du \int \hat{\eta}_k(u) f(u, x) du} \right\} dz \leq \|\varphi\| C\epsilon$$

where, again (A1) (1) has been applied along with Theorem 2 of [22] and  $C$  does not depend upon  $x, k, \epsilon$  or  $y_{0:n}$ . Using the uniformity in  $x$  of the above bounds allows us to conclude.  $\square$

## References

- [1] ANDRIEU, C. & ROBERTS, G. O. (2009). The pseudo-marginal approach for efficient Monte Carlo computations. *Ann. Statist.*, **37**, 697–725.
- [2] ANDRIEU, C., DOUCET, A. & HOLENSTEIN, R. (2010). Particle Markov chain Monte Carlo methods (with discussion). *J. R. Statist. Soc. Ser. B*, **72**, 269–342.
- [3] ANDRIEU, C., DOUCET, A. & GODSILL, S. (2000). On sequential Monte Carlo sampling methods for Bayesian filtering. *Statist. Comp.*, **10**, 197–208.
- [4] BARTHELMÉ, S. & CHOPIN, N. (2011). Expectation-Propagation for Summary-Less, Likelihood-Free Inference, Technical Report, ENSAE.
- [5] BESKOS, A., CRISAN, D., JASRA, A. & WHITELEY, N. P. (2011). Error and normalizing constants for sequential Monte Carlo methods in high-dimensions. Technical Report, Imperial College London.
- [6] BRIERS, M. & DOUCET, A. & MASKELL, S. (2010). Smoothing algorithms for state-space models. *Ann. Inst. Statist. Math.*, **62**, 61–89.
- [7] CALVET, C. & CZELLAR, V. (2012). Accurate Methods for Approximate Bayesian Computation Filtering. Technical Report, HEC Paris.
- [8] CAMPILLO, F., & ROSSI, V. (2009). Convolution Particle Filter for Parameter Estimation in General State-Space Models. *IEEE Trans. Aero. Elec. Sys.*, **45**, 1063–1072.
- [9] CAPPÉ, O., RYDEN, T., & MOULINES, É. (2005). *Inference in Hidden Markov Models*. Springer: New York.
- [10] CHOPIN, N., JACOB, P. & PAPASPILIOPOULOS, O. (2011). SMC<sup>2</sup>: A sequential Monte Carlo algorithm with particle Markov chain Monte Carlo updates. Technical Report. ENSAE.
- [11] CORNEBISE, J. & PETERS, G. (2010). Discussion of particle Markov chain Monte Carlo, *J. R. Statist. Soc. Ser. B*, **72**, 329–330.
- [12] DEAN, T., SINGH, S. S., JASRA, A., & PETERS, G. (2010). Parameter inference for hidden Markov models with intractable likelihoods. Technical Report, University of Cambridge.
- [13] DEAN, T. & SINGH, S.S. (2011) Asymptotic behaviour of approximate Bayesian estimators Technical Report, University of Cambridge.
- [14] DEL MORAL, P. (2004). *Feynman-Kac Formulae: Genealogical and Interacting Particle Systems with Applications*. Springer: New York.
- [15] DEL MORAL, P., DOUCET, A. & SINGH, S. S. (2010). A backward interpretation of Feynman-Kac formulae. *M2AN*, **44**, 947–975.
- [16] DEL MORAL, P., DOUCET, A. & SINGH, S. S. (2010). Forward smoothing using sequential Monte Carlo. Technical Report, University of Cambridge.

- [17] DOUC, R., GARIVIER, A., MOULINES, E. & OLSSON, J. (2011). Sequential Monte Carlo smoothing for general state space Hidden Markov Models. *Ann. Appl. Probab.*, **21**, 2109–2145.
- [18] DOUCET, A. & JOHANSEN, A. M. (2011). Particle filtering and smoothing: Fifteen years later. In *Oxford Handbook of Nonlinear Filtering*, D. Crisan and B. Rozovsky (eds.).OUP: Oxford.
- [19] EHRLICH, E., JASRA, A., KANTAS, N. & SINGH, S. S. (2012). Online Static Parameter Estimation for ABC Approximation of Hidden Markov Models. Technical Report, Imperial College London (to be submitted).
- [20] GAUCHI, J.-P. & VILA, J.-P. (2012). Nonparametric filtering approaches for identification and inference in nonlinear dynamic systems. *Statist. Comp.* (to appear).
- [21] GODSILL, S. J., DOUCET, A. & WEST, M. (2004). Monte Carlo smoothing for nonlinear time series. *J. Amer. Stat. Assoc.*, **99**, 156-168.
- [22] JASRA, A., SINGH, S., MARTIN, J., & MCCOY, E. (2012). Filtering via Approximate Bayesian computation. *Statist. Comp* (to appear).
- [23] LIU, J.S. (2001) *Monte Carlo Strategies in Scientific Computing*. New York: Springer Verlag.
- [24] MARIN, J.-M., PUDLO, P., ROBERT, C.P., & RYDER, R. (2012). Approximate Bayesian Computational methods. *Statist. Comp.* (to appear).
- [25] MCKINLEY, J., COOK, A., & DEARDON, R. (2009). Inference for epidemic models without likelihoods. *Intl. J. Biostat.*, **5**.
- [26] OLSSON, J, & RYDEN, T. (2011). Rao-Blackwellisation of particle Markov chain Monte Carlo methods using forward filtering backward sampling. *IEEE Trans. Sig. Proc.*, **59**, 4606–4619.
- [27] POYIADJIS, G., DOUCET, A. & SINGH, S.S. (2011). Particle approximations of the score and observed information matrix in state-space models with application to parameter estimation. *Biometrika*, **98**, 65–80.
- [28] YILDIRIM, S., SINGH, S.S. & JASRA, A. (2012). Expectation Maximisation for approximate Bayesian computation maximum Likelihood estimation. Technical Report, University of Cambridge (to be submitted).