

RESEARCH

Open Access



Online sequential Monte Carlo smoother for partially observed diffusion processes

Pierre Gloaguen¹, Marie-Pierre Étienne¹ and Sylvain Le Corff^{2*}

Abstract

This paper introduces a new algorithm to approximate smoothed additive functionals of partially observed diffusion processes. This method relies on a new sequential Monte Carlo method which allows to compute such approximations online, i.e., as the observations are received, and with a computational complexity growing linearly with the number of Monte Carlo samples. The original algorithm cannot be used in the case of partially observed stochastic differential equations since the transition density of the latent data is usually unknown. We prove that it may be extended to partially observed continuous processes by replacing this unknown quantity by an unbiased estimator obtained for instance using general Poisson estimators. This estimator is proved to be consistent and its performance are illustrated using data from two models.

Keywords: Stochastic differential equations, Smoothing, Sequential Monte Carlo Methods, Online estimation

1 Introduction

This paper introduces a new algorithm to solve the smoothing problem for hidden Markov models (HMMs) whose hidden state is a solution to a stochastic differential equation (SDE). These models are referred to as partially observed diffusion (POD) processes in [27]. The hidden state process $(X_t)_{t \geq 0}$ is assumed to be a solution to a SDE, and the only information available is given by noisy observations $(Y_k)_{0 \leq k \leq n}$ of the states $(X_k)_{0 \leq k \leq n}$ (where X_k stands for X_{t_k}) at some discrete time points $(t_k)_{0 \leq k \leq n}$. The bivariate stochastic process $\{(X_k, Y_k)\}_{0 \leq k \leq n}$ is a state-space model such that conditional on the state sequence $(X_k)_{0 \leq k \leq n}$ the observations $(Y_k)_{0 \leq k \leq n}$ are independent and for all $0 \leq \ell \leq n$ the conditional distribution of Y_ℓ given $\{X_k\}_{0 \leq k \leq n}$ depends on X_ℓ only.

Statistical inference for HMMs often requires to solve Bayesian filtering and smoothing problems, i.e., the computation of the posterior distributions of sequences of hidden states given observations. The filtering problem refers to the estimation, for each $0 \leq k \leq n$, of the distributions of the hidden state X_k given the observations (Y_0, \dots, Y_k) . Smoothing stands for the estimation of the

distributions of the sequence of states (X_k, \dots, X_p) given observations (Y_0, \dots, Y_ℓ) with $0 \leq k \leq p \leq \ell \leq n$. These posterior distributions are crucial to compute maximum likelihood estimators of unknown parameters using the observations (Y_0, \dots, Y_n) only. For instance, the E-step of the EM algorithm introduced in [9] boils down to the computation of a conditional expectation of an additive functional of the hidden states given all the observations up to time n . Similarly, by Fisher's identity, recursive maximum likelihood estimates may be computed using the gradient of the log likelihood which can be written as the conditional expectation of an additive functional of the hidden states. See [7, Chapters 10 and 11], [19, 23, 24, 31] for further references on the use of these smoothed expectations of additive functionals applied to maximum likelihood parameter inference in latent data models.

However, in most cases, the exact computation of these expectations is usually not possible explicitly. Sequential Monte Carlo (SMC) methods are popular algorithms to approximate smoothing distributions with random particles associated with importance weights. [17, 22] introduced the first particle filters and smoothers for state-space models by combining importance sampling steps to propagate particles with resampling steps to duplicate or discard particles according to their importance weights. In the case of HMMs, approximations of the smoothing distributions may be obtained using the

*Correspondence: sylvain.lecorff@math.u-psud.fr

²Laboratoire de Mathématiques d'Orsay, Univ. Paris-Sud, CNRS, Université Paris-Saclay, 91405 Orsay, France

Full list of author information is available at the end of the article

forward filtering backward smoothing algorithm (FFBS) and the forward filtering backward simulation algorithm (FFBSi) developed respectively in [11, 18, 22], and [16]. Both algorithms require first a forward pass which produces a set of particles and weights approximating the sequence of filtering distributions up to time n . Then, a backward pass is performed to compute new weights (FFBS) or sample trajectories (FFBSi) in order to approximate the smoothing distributions. Recently, [28] proposed a new SMC algorithm, the particle-based rapid incremental smoother (PaRIS), to approximate on-the-fly (i.e., using the observations as they are received) smoothed expectations of additive functionals. Unlike the FFBS algorithm, the complexity of this algorithm grows only linearly with the number of particles N and contrary to the FFBSi algorithm, no backward pass is required. One of the best features of PaRIS algorithm is that it may be implemented online, using the observations $(Y_k)_{k \geq 0}$ as they are received, without any increasing storage requirements.

Unfortunately, these methods cannot be applied directly to POD processes since some elementary quantities, such as transition densities of the hidden states, are not available explicitly. In the context of SDEs, discretization procedures may be used to approximate transition densities. For instance, the classical Euler-Maruyama method, the Ozaki discretization which proposes a linear approximation of the drift coefficient between two observations [29, 32], or Gaussian-based approximations using Taylor expansions of the conditional mean and variance of an observation given the observation at the previous time step, [20, 21, 33]. Other approaches based on Hermite polynomials expansion were also introduced by [1–3] and were extended in several directions recently, see [25] and all the references on the approximation of transition densities therein. However, even the most recent discretization-based approximations of the transition densities induce a systematic bias in the approximation of the transition densities, see for instance [8].

To overcome this difficulty, [13] proposed to solve the filtering problem by combining SMC methods with an unbiased estimate of the transition densities based on the generalized Poisson estimator (GPE). In this case, only the Monte Carlo error has to be controlled as there is no Taylor expansion to approximate unknown transition densities, i.e., no discretization scheme is used. The only solution to solve the smoothing problem for POD processes using SMC methods without any discretization procedure has been proposed in [27] and extends the fixed-lag smoother of [26]. Using forgetting properties of the hidden chain, the algorithm improves the performance of [13] to approximate smoothing distributions but at the cost of a bias, this time due to the fixed lag approximation, that does not vanish as the number of particles grows to infinity.

In this paper, we propose to use SMC methods to obtain consistent approximations of smoothing expectations of POD processes by extending the PaRIS algorithm. The proposed algorithm allows to approximate smoothed expectations of additive functionals online, with a complexity growing only linearly with the number of particles and without any discretization procedure or Taylor expansion of the transition densities. The crucial and simple result (Lemma 1) of the application of the PaRIS algorithm to POD processes is that the acceptance rejection mechanism introduced in [10] ensuring the linear complexity of the procedure is still correct when the transition densities are replaced by unbiased estimates. The usual FFBS and FFBSi algorithms may not extend this easily since they both require the computation of weights defined as ratios involving the transition densities, thus replacing these unknown quantities by unbiased estimates does not lead to unbiased estimators of the weights. The linear version of the FFBSi algorithm proposed in [10] could be extended in a similar way as PaRIS algorithm but it would still require a backward pass and would not be an online smoother. The proposed generalized random version of PaRIS algorithm, hereafter named GRandom PaRIS algorithm, may not only be applied to POD processes but also to any general state-space model where the transition density of the hidden chain may be approximated using a positive and unbiased estimator.

Section 2 describes the model and the smoothing quantities to be estimated. Section 3 provides the algorithm to approximate smoothed additive functionals using unbiased estimates of the transition density of the hidden states. This section also details the application of this algorithm when the transition density are approximated using a GPE. In Section 4, classical convergence results for SMC smoothers are extended to the setting of this paper and illustrated with numerical experiments in Section 5. All proofs are postponed to Appendix.

2 Model and framework

Let $(X_t)_{t \geq 0}$ be defined as a weak solution to the following SDE in \mathbb{R}^d :

$$X_0 = x_0 \quad \text{and} \quad dX_t = \alpha(X_t)dt + \Gamma(X_t)dW_t, \quad (1)$$

where $(W_t)_{t \geq 0}$ is a standard Brownian motion on \mathbb{R}^d , $\alpha : \mathbb{R}^d \rightarrow \mathbb{R}^d$, and $\Gamma : \mathbb{R}^d \rightarrow \mathbb{R}^{d \times d}$. The solution to (1) is supposed to be partially observed at times $t_0 = 0, \dots, t_n$ through an observation process $(Y_k)_{0 \leq k \leq n}$ in $(\mathbb{R}^m)^{n+1}$. In the following, for all $0 \leq k \leq n$, the state X_{t_k} at time k is referred to as X_k . For all $0 \leq k \leq n$, the distribution of Y_k given X_k has a density with respect to a reference measure λ on \mathbb{R}^m given by $g(X_k, \cdot)$. For the sake of simplicity, the shorthand notation $g_k(X_k)$ for $g(X_k, Y_k)$ is used. The distribution of X_0 has a density with respect to a reference measure μ on \mathbb{R}^d given by χ . For all $0 \leq k \leq n - 1$,

the conditional distribution of X_{k+1} given X_k has a density $q_k(X_k, \cdot)$ with respect to μ .

For all $0 \leq k \leq k' \leq n$, the joint smoothing distributions of the hidden states are defined, for all measurable function h on $(\mathbb{R}^d)^{k'-k+1}$, by:

$$\phi_{k:k'|n}[h] = \mathbb{E}[h(X_k, \dots, X_{k'}) | Y_{0:n}]$$

and $\phi_k = \phi_{k:k|k}$ denotes the filtering distributions. The aim of this paper is to approximate expectations of the form:

$$\begin{aligned} \phi_{0:n|n}[H_n] &= \mathbb{E}[H_n(X_{0:n}) | Y_{0:n}], \\ \text{where } H_n &= \sum_{k=0}^{n-1} h_k(X_k, X_{k+1}), \end{aligned} \tag{2}$$

when $\{h_k\}_{k=0}^{n-1}$ are given functions on $\mathbb{R}^d \times \mathbb{R}^d$. Smoothed additive functionals as (2) are crucial for maximum likelihood inference of latent data models. These quantities appear naturally when computing the Fisher score in hidden Markov models or the intermediate quantity of the expectation maximization algorithm (see Section 5). They are also pivotal to design online expectation maximization-based algorithms which motivates the method introduced in this paper that does not require growing storage and can process observations online.

The algorithm proposed in this paper is based on sequential Monte Carlo methods which offer a flexible framework to approximate such distributions with weighted empirical measures associated with random samples. At each time step, the samples are moved randomly in \mathbb{R}^d and associated with importance weights. In general situations, the computation of these importance weights involve the unknown transition density of the process (1). The solution introduced in Section 3 requires an unbiased estimator of these unknown transition densities. Moreover, this estimator must be almost surely positive and upper bounded. Statistical inference of stochastic differential equations is an active area of research, and several solutions have been proposed to design unbiased estimates of these transition densities. Those estimators require different assumptions on the model (1), we provide below several solutions that can be investigated.

General Poisson estimators This paper focuses mainly on GPEs which have been widely used recently and applied in a variety of disciplines. These estimators require that the diffusion coefficient Γ is constant and equal to the identity matrix, see [13]. They may be applied to reducible SDE for which there exists an invertible and infinitely differentiable function η such that the process $\{Z_t = \eta(X_t)\}_{t \geq 0}$ satisfies the SDE $z_0 = \eta(x_0)$ and

$$dZ_t = \beta(Z_t)dt + dW_t. \tag{3}$$

By Ito's formula, it is straightforward to show that, in the case of a reducible diffusion, the Jacobian matrix of η satisfies

$$\nabla \eta = \Gamma^{-1},$$

and, in the case $d = 1$,

$$\beta : u \mapsto \frac{\alpha(\eta^{-1}(u))}{\Gamma(\eta^{-1}(u))} - \frac{\Gamma'(\eta^{-1}(u))}{2}.$$

In the case of a scalar diffusion, this Lamperti transform is given by

$$\eta : x \mapsto \int_{x_0}^x \Gamma^{-1}(u)du.$$

In the general case, [3, Proposition 1] shows that when Γ is non-singular, the SDE is reducible if and only if, for all $1 \leq i, j, k \leq d$,

$$\frac{\partial \Gamma_{ij}^{-1}}{\partial x_k} = \frac{\partial \Gamma_{ik}^{-1}}{\partial x_j}. \tag{4}$$

In the case of a diagonal matrix Γ (4) is equivalent to assume that Γ is such that for each $1 \leq i \leq d$, $\Gamma_{i,i}$ depends on x_i only. [3] notes that the reducibility condition (4) holds also for some non-diagonal matrices Γ . This is true in particular in the case $d = 2$ for stochastic volatility models where σ is of the form:

$$\Gamma(x) = \begin{pmatrix} a(x_1) & a(x_1)b(x_2) \\ 0 & c(x_2) \end{pmatrix}.$$

GPEs consider that the process $(X_t)_{t \geq 0}$ satisfies the SDE (1) with Γ being the identity matrix, i.e., we consider a diffusion after the application of the Lamperti transform. In addition, designing GPEs also requires that

- i) α is of the form $\alpha(x) = \nabla_x A(x)$ where $A : \mathbb{R}^d \rightarrow \mathbb{R}$ is a twice continuously differentiable function ;
- ii) the function $x \mapsto (\|\alpha(x)\|^2 + \Delta A(x)) / 2$ is lower bounded where Δ is the Laplace operator.

Assumption (i) is somewhat restrictive as it requires α to derive from a scalar potential, however, it has natural applications in many fields such as movement ecology, see [15]. Assumption (ii) is a technical condition which ensures that exact sampling of processes solution to (1) using acceptance rejection methods, see for instance [4, 5, 13]. In addition to provide an unbiased estimate of the transition density, the GPE ensure that this estimate is almost surely positive. Moreover, as detailed below, under additional conditions, a GPE that is almost surely upper bounded can be defined.

Continuous importance sampling-based estimators

In the case the previous assumptions are not fulfilled, in particular assumption (i), alternatives to GPEs are given by continuous importance sampling procedures for SDE. In

[34], for each $0 \leq k \leq n-1$, the transition density between t_k and t_{k+1} is expressed as an infinite expansion obtained using the Kolmogorov backward operator associated with (1). This analytical expression of the transition density is not tractable and is estimated by updating random samples at random times between t_k and t_{k+1} using tractable proposal distributions (for instance, based on an Euler discretization of the original SDE). Then, these samples are associated with random weights to ensure that the proposed estimator is unbiased. More recently, [14] extended the discrete time importance sampling estimator by introducing updates at random times associated with a renewal process. The random samples are weighted using the Kolmogorov forward operator associated with the SDE which relies on the first two order derivatives of the drift and diffusion coefficients (and is therefore tractable).

The unbiasedness of these procedures and the controls of the variability of the estimates require moments assumptions and Holder type conditions on the parameters of the SDE (1). Their efficiency require a fair amount of tuning as they highly depend on the proposal densities used to obtain the Monte Carlo samples and the point processes generating the underlying random times. In addition to unbiasedness, the proposed algorithm in this work requires that the estimator of the transition density is almost surely positive and upper bounded. This implies additional assumptions on the SDE depending on the chosen estimate and could lead to interesting perspectives.

3 The generalized random PaRIS algorithm

The algorithm is based on the following link between the filtering and smoothing distributions for additive functionals, see [28]:

$$\begin{aligned} \phi_{0:n|n}[h] &= \phi_n[T_n[h]], \text{ where} \\ T_n[h](X_n) &= \mathbb{E}[h(X_{0:n})|X_n, Y_{0:n}]. \end{aligned} \tag{5}$$

The approximation of (5) requires first to approximate the sequence of filtering distributions. Sequential Monte Carlo methods provide an efficient and simple solution to obtain these approximations using sets of particles $\{\xi_k^\ell\}_{\ell=1}^N$ associated with weights $\{\omega_k^\ell\}_{\ell=1}^N$, $0 \leq k \leq n$.

At time $k = 0$, N particles $\{\xi_0^\ell\}_{\ell=1}^N$ are sampled independently according to $\xi_0^\ell \sim \eta_0$, where η_0 is a probability density with respect to μ . Then, ξ_0^ℓ is associated with the importance weights $\omega_0^\ell = \chi(\xi_0^\ell) g_0(\xi_0^\ell) / \eta_0(\xi_0^\ell)$. For any bounded and measurable function h defined on \mathbb{R}^d , the expectation $\phi_0[h]$ is approximated by

$$\phi_0^N[h] = \frac{1}{\Omega_0^N} \sum_{\ell=1}^N \omega_0^\ell h(\xi_0^\ell), \quad \Omega_0^N := \sum_{\ell=1}^N \omega_0^\ell.$$

Then, for $1 \leq k \leq n$, using $\left\{ \left(\xi_{k-1}^\ell, \omega_{k-1}^\ell \right) \right\}_{\ell=1}^N$, the auxiliary particle filter of [30] samples pairs $\left\{ \left(I_k^\ell, \xi_k^\ell \right) \right\}_{\ell=1}^N$ of indices and particles using an instrumental transition density p_{k-1} on $\mathbb{R}^d \times \mathbb{R}^d$ and an adjustment multiplier function ϑ_k on \mathbb{R}^d . Each new particle ξ_k^ℓ and weight ω_k^ℓ at time k are computed following these steps:

- choose a particle index I_k^ℓ at time $k - 1$ in $\{1, \dots, N\}$ with probabilities proportional to $\omega_{k-1}^j \vartheta_k(\xi_{k-1}^j)$, for j in $\{1, \dots, N\}$;
- sample ξ_k^ℓ using this chosen particle according to $\xi_k^\ell \sim p_{k-1}(\xi_{k-1}^{I_k^\ell}, \cdot)$;
- associate the particle ξ_k^ℓ with the importance weight:

$$\omega_k^\ell := \frac{q_{k-1}(\xi_{k-1}^{I_k^\ell}, \xi_k^\ell) g_k(\xi_k^\ell)}{\vartheta_k(\xi_{k-1}^{I_k^\ell}) p_{k-1}(\xi_{k-1}^{I_k^\ell}, \xi_k^\ell)}. \tag{6}$$

The expectation $\phi_k[h]$ is approximated by

$$\phi_k^N[h] := \frac{1}{\Omega_k^N} \sum_{\ell=1}^N \omega_k^\ell h(\xi_k^\ell), \quad \Omega_k^N := \sum_{\ell=1}^N \omega_k^\ell.$$

The most simple choice for p_{k-1} and ϑ_k is the bootstrap filter proposed by [17] which sets $p_{k-1} = q_{k-1}$ and for all $x \in \mathbb{R}^d$, $\vartheta(x) = 1$. In the case of POD processes, q_{k-1} is unknown but it can be replaced by any approximation to sample the particles as any choice of p_{k-1} can be made. The approximation can be obtained using a discretization scheme such as Euler method or a Poisson-based approximation as detailed below. A more appealing choice is the fully adapted particle filter which sets for all $x, x' \in \mathbb{R}^d$, $p_{k-1}(x, x') \propto q_{k-1}(x, x') g_k(x')$ and for all $x \in \mathbb{R}^d$, $\vartheta(x) = \int q_{k-1}(x, x') g_k(x') \mu(dx')$. Here, again q_{k-1} has to be replaced by an approximation. In Section 5, it is replaced by the Gaussian approximation provided by a Euler scheme which leads to a Gaussian proposal density p_{k-1} as the observation model is linear and Gaussian.

The PaRIS algorithm uses the same decomposition as the FFBS algorithm introduced in [12] and the FFBSi algorithm proposed by [16] to approximate smoothing distributions. It combines both the forward-only version of the FFBS algorithm with the sampling mechanism of the FFBSi algorithm. It does not produce an approximation of the smoothing distributions but of the smoothed expectation of a fixed additive functional and thus may be used to approximate (2). Its crucial property is that it does not require a backward pass, the smoothed expectation is computed on-the-fly with the particle filter and no storage of the particles or weights is needed.

PaRIS algorithm relies on the following fundamental property of $T_k[H_k]$ when H_k is as in (2):

$$T_k[H_k](X_k) = \mathbb{E} \left[T_{k-1}[H_{k-1}](X_{k-1}) + h_{k-1}(X_{k-1}, X_k) \middle| X_k, Y_{0:k-1} \right],$$

$$= \frac{\int \phi_{k-1}(dx_{k-1}) q_{k-1}(x_{k-1}, X_k) \{ T_{k-1}[H_{k-1}](x_{k-1}) + h_{k-1}(x_{k-1}, X_k) \}}{\int \phi_{k-1}(dx_{k-1}) q_{k-1}(x_{k-1}, X_k)}.$$

Therefore, [28] introduces sufficient statistics τ_k^i (starting with $\tau_0^i = 0, 1 \leq i \leq N$), approximating $T_k[H_k](\xi_k^i)$, for $1 \leq i \leq N$ and $0 \leq k \leq n$. First, replacing ϕ_{k-1} by ϕ_{k-1}^N in the last equation leads to the following approximation of $T_k[H_k](\xi_k^i)$:

$$T_k^N[H_k](\xi_k^i) = \sum_{j=1}^N \Lambda_{k-1}^N(i, j) \left\{ T_{k-1}[H_{k-1}](\xi_{k-1}^j) + h_{k-1}(\xi_{k-1}^j, \xi_k^i) \right\}, \quad (7)$$

where

$$\Lambda_k^N(i, \ell) = \frac{\omega_k^\ell q_k(\xi_k^\ell, \xi_{k+1}^i)}{\sum_{\ell=1}^N \omega_k^\ell q_k(\xi_k^\ell, \xi_{k+1}^i)}, \quad 1 \leq \ell \leq N. \quad (8)$$

Computing exactly these approximations would lead to a complexity growing quadratically with N because of the normalizing constant in (8). Therefore, PaRIS algorithm sample particles in the set $\{\xi_{k-1}^j\}_{j=1}^N$ with probabilities $\Lambda_k^N(i, \cdot)$ to approximate the expectation (7) and produce τ_k^i . Choosing $\tilde{N} \geq 1$, at each time step $0 \leq k \leq n-1$ these statistics are updated according to the following steps:

- (i) Run one step of a particle filter to produce $\{(\xi_k^\ell, \omega_k^\ell)\}$ for $1 \leq \ell \leq N$.
- (ii) For all $1 \leq i \leq N$, sample independently $J_k^{i,\ell}$ in $\{1, \dots, N\}$ for $1 \leq \ell \leq \tilde{N}$ with probabilities $\Lambda_k^N(i, \cdot)$, given by (8).
- (iii) Set

$$\tau_{k+1}^i := \frac{1}{\tilde{N}} \sum_{\ell=1}^{\tilde{N}} \left\{ \tau_k^{J_k^{i,\ell}} + h_k \left(\xi_k^{J_k^{i,\ell}}, \xi_{k+1}^i \right) \right\}.$$

Then, (2) is approximated by

$$\phi_{0:n|n}^N[\tau_n] = \frac{1}{\Omega_n^N} \sum_{i=1}^N \omega_n^i \tau_n^i.$$

It is clear from steps (i) to (iii) that each time a new observation Y_{n+1} is received, the quantities $(\tau_{n+1}^i)_{1 \leq i \leq N}$ can be updated only using Y_{n+1} , $(\tau_n^i)_{1 \leq i \leq N}$ and the particle filter at time n . This means that storage requirements do not increase when processing additional data.

As proved in [28], the algorithm is asymptotically consistent (as N goes to infinity) for any precision parameter \tilde{N} . However, there is a significant qualitative difference

between the cases $\tilde{N} = 1$ and $\tilde{N} \geq 2$. As for the FFBSI algorithm, when there exists σ_+ such that $0 < q_k < \sigma_+$, PaRIS algorithm may be implemented with $\mathcal{O}(N)$ complexity using the accept-reject mechanism of [10].

In general situations, PaRIS algorithm cannot be used for stochastic differential equations as q_k is unknown. Therefore, the computation of the importance weights ω_k^ℓ and of the acceptance ratio of [10] is not tractable. Following [13, 27], filtering weights can be approximated by replacing $q_k(\xi_k^\ell, \xi_{k+1}^i)$ by an unbiased estimator $\hat{q}_k(\xi_k^\ell, \xi_{k+1}^i; \zeta_k)$, where ζ_k is a random variable in \mathbb{R}^q such that

$$\hat{q}_k(\xi_k^\ell, \xi_{k+1}^i; \zeta_k) > 0 \text{ a.s. and}$$

$$\mathbb{E} \left[\hat{q}_k(\xi_k^\ell, \xi_{k+1}^i; \zeta_k) \middle| \mathcal{G}_{k+1}^N \right] = q_k(\xi_k^\ell, \xi_{k+1}^i),$$

where for all $0 \leq k \leq n$,

$$\mathcal{F}_k^N = \sigma \left\{ Y_{0:k}; \left(\xi_u^\ell, \omega_u^\ell, \tau_u^\ell \right); J_v^{\ell,j}; 1 \leq \ell \leq N, 0 \leq u \leq k, 1 \leq j \leq \tilde{N}, 0 \leq v < k \right\},$$

$$\mathcal{G}_{k+1}^N = \mathcal{F}_k^N \vee \sigma \left\{ Y_{k+1}; \left(\xi_{k+1}^\ell, \omega_{k+1}^\ell \right); 1 \leq \ell \leq N \right\}.$$

Practical choices for ζ_k are discussed below, see for instance (14) which presents the choice made for the implementation of such estimators in our context. In the case where q_k is unknown, the filtering weights in (6) then become

$$\hat{\omega}_k^\ell := \frac{\hat{q}_{k-1} \left(\xi_{k-1}^{\ell}, \xi_k^\ell; \zeta_{k-1} \right) g_k \left(\xi_k^\ell \right)}{\vartheta_k \left(\xi_{k-1}^{\ell} \right) p_{k-1} \left(\xi_{k-1}^{\ell}, \xi_k^\ell \right)}. \quad (9)$$

In Algorithm 1, M independent copies $(\zeta_{k-1}^m)_{1 \leq m \leq M}$ of ζ_{k-1} are sampled and the empirical mean of the associated estimates of the transition density are used to compute $\hat{\omega}_k^\ell$ instead of a single realization. Therefore, to obtain a generalized random version of PaRIS algorithm, we only need to be able to sample from the discrete probability distribution $\Lambda_k^N(i, \cdot)$ in the case of POD processes.

Consider the following assumption: for all $0 \leq k \leq n-1$, there exists a random variable $\hat{\sigma}_+^k$ measurable with respect to \mathcal{G}_{k+1}^N such that,

$$\sup_{x,y,\zeta} \hat{q}_k(x, y; \zeta) \leq \hat{\sigma}_+^k. \quad (10)$$

Lemma 1 Assume that (10) holds for some $0 \leq k \leq n-1$. For all $1 \leq i \leq N$, define the random variable J_k^i as follows:

repeat

Sample independently $\zeta, U \sim \mathcal{U}[0, 1]$, and $J \in \{1, \dots, N\}$ with probabilities proportional to $\{\hat{\omega}_k^1, \dots, \hat{\omega}_k^N\}$.

until $U \leq \widehat{q}_k \left(\xi_k^J, \xi_{k+1}^i, \zeta \right) / \widehat{\sigma}_+^k$.

Set $J_k^i = J$.

Then, the conditional probability distribution given \mathcal{G}_{k+1}^N of J_k^i is $\Lambda_k^N(i, \cdot)$.

Proof See [Appendix](#). □

Note that Lemma 1 still holds if assumption (10) is relaxed and replaced by

$$\sup_{j,y,\zeta} \widehat{q}_k \left(\xi_k^j, y, \zeta \right) \leq \widehat{\sigma}_+^k. \tag{11}$$

It is worth noting that under assumptions (10) or (11), the linear complexity property of PaRIS algorithm is ensured. The following assumption can also be considered. For all $1 \leq i \leq N$,

$$\sup_{j,\zeta} \widehat{q}_k \left(\xi_k^j, \xi_{k+1}^i, \zeta \right) \leq \widehat{\sigma}_+^{k,i}. \tag{12}$$

If only assumption (12) holds, the algorithm has a quadratic complexity. The bound of (10) is uniform (it does not depend on the particles) and can be used for every particle $1 \leq i \leq N$. However, this bound can be large (with respect to the simulated set of particles) for the algorithm of Lemma 1. The bound of (12) requires N computations per particle (therefore, N^2 computations). However, it is clear that this second bound is sharper than the one of (10) for the acceptance rejection procedure and may lead to a computationally more efficient algorithm.

Bounded estimator of q_k using GPEs For $x, y \in \mathbb{R}^d$, by Girsanov and Ito’s formulas, the transition density $q_k(x, y)$ of (1) satisfies, with $\Delta_k = t_{k+1} - t_k$,

$$q_k(x, y) = \varphi_{\Delta_k}(x, y) \exp \{A(y) - A(x)\} \mathbb{E}_{\mathbb{W}^{x,y,\Delta_k}} \left[\exp \left\{ - \int_0^{\Delta_k} \phi(w_s) ds \right\} \right],$$

where $\mathbb{W}^{x,y,\Delta_k}$ is the law of Brownian bridge starting at x at 0 and hitting y at Δ_k , $(w_t)_{0 \leq t \leq \Delta_k}$ is such a Brownian bridge, $\varphi_{\Delta_k}(x, y)$ is the p.d.f. of a normal distribution with mean x and variance Δ_k , evaluated at y and $\phi : \mathbb{R}^d \rightarrow \mathbb{R}$ is defined as

$$\phi(x) = (\|\alpha(x)\|^2 + \Delta A(x)) / 2.$$

Assume that there exist random variables L_w and U_w such that for all $0 \leq s \leq \Delta_k$, $L_w \leq \phi(w_s) \leq U_w$. The performance of the estimator depends on the choice of L_w and U_w which is specific to the SDE. In the case of the models analyzed in Section 5, these bounds are discussed in [13] for the SINE model and in [27] for the log-growth model. Note that in the case where ϕ is not

upper bounded, [5] proposed the EA3 algorithm. This *layered Brownian bridge* construction first samples random variables to determine in which *layer* the Brownian bridge lies before simulating the bridge conditional on the event that it belongs to the layer. By continuity of ϕ , L_w , and U_w can be computed easily.

Let κ be a random variable taking values in \mathbb{N} with distribution μ and $(U_j)_{1 \leq j \leq \kappa}$ be independent uniform random variables on $[0, \Delta_k]$ and $\zeta_k = \{\kappa, w, U_1, \dots, U_\kappa\}$. As shown in [13], a positive unbiased estimator is given by

$$\widehat{q}_k(x, y; \zeta_k) = \varphi_{\Delta_k}(x, y) \exp \{A(y) - A(x)\} \times \exp \{-U_w \Delta\} \frac{\Delta_k^\kappa}{\mu(\kappa)\kappa!} \prod_{j=1}^{\kappa} (U_w - \phi(w_{U_j})). \tag{13}$$

Interesting choices of μ are discussed in [13], and we focus here on the so called GPE-1, where μ is a Poisson distribution with intensity $(U_w - L_w)\Delta_k$. In that case, the estimator (13) becomes

$$\widehat{q}_k(x, y; \zeta_k) = \varphi_{\Delta_k}(x, y) \exp \{A(y) - A(x) - L_w \Delta_k\} \prod_{j=1}^{\kappa} \frac{U_w - \phi(w_{U_j})}{U_w - L_w}. \tag{14}$$

On the r.h.s. of (14), the product over κ elements is bounded by 1. Therefore, a sufficient condition to satisfy one of the assumptions (10)–(12) is that the function

$$\rho_{\Delta_k} : \mathbb{R}^d \times \mathbb{R}^d \mapsto \mathbb{R} \quad (x, y) \mapsto \varphi_{\Delta_k}(x, y) \exp \{A(y) - A(x) - L_w \Delta_k\}$$

is upper bounded almost surely by $\widehat{\sigma}_+^k$. In particular, if L_w is bounded below almost surely, (14) always satisfies assumption (12) and Algorithm 1 can be used. This condition is always satisfied for models in the domains required for the applications of exact algorithms EA1, EA2, and EA3 defined in [6].

When (10) or (11) holds, it can be nonetheless of practical interest to choose the bounds $\widehat{\sigma}_+^{k,i}$, $1 \leq i \leq N$, corresponding to (12). Indeed, this might increase significantly the acceptance rate of the algorithm, and therefore reduce the number of draws of the random variable ζ_k , which has a much higher cost than the computation of ρ_{Δ_k} , as it requires simulations of Brownian bridges. Moreover, this option allows to avoid numerical optimization if no analytical expression of $\widehat{\sigma}_+^k$ is available. In practice, this seems more efficient in terms of computational time when N has moderate values.

Algorithm 1 GRand PaRIS algorithm

for all $i \in 1, \dots, N$ **do**
 Sample $\xi_0^i \sim \eta_0$, $\tau_0^i = 0$ and $\widehat{\omega}_0^i = g_0(\xi_0^i) \chi_0(\xi_0^i) / \eta_0(\xi_0^i)$.
end for
for $k \in 0, \dots, n - 1$ **do**
for all $i \in 1, \dots, N$ **do**
 Set $\tau_{k+1}^i = 0$;
 Sample I_{k+1}^i in $\{1, \dots, N\}$ with probabilities proportional to $\{\widehat{\omega}_k^1 \vartheta_{k+1}(\xi_k^1), \dots, \widehat{\omega}_k^N \vartheta_{k+1}(\xi_k^N)\}$.
 Sample $\xi_{k+1}^i \sim p_k(\xi_k^{I_{k+1}^i}, \cdot)$.
 For all $1 \leq m \leq M$, sample independently $\zeta_k^m = (\kappa_m, \mathbf{w}_m, (U_j^m)_{1 \leq j \leq \kappa_m})$ with $\kappa_m \sim \mu$, $\mathbf{w}_m \sim \mathbb{W}_k^{X_{k+1}}$ and $(U_j^m)_{1 \leq j \leq \kappa_m} \sim \mathcal{U}[0, \Delta_k]^{\otimes \kappa_m}$.
 Compute $\widehat{\omega}_{k+1}^i$ using Eq. (9).
for all $\ell \in 1, \dots, \tilde{N}$ **do**
 Sample $J_k^{i,\ell}$ as in Lemma 1.
 Update $\tau_{k+1}^i = \tau_{k+1}^i + (\tau_k^{j,\ell} + h_k(\xi_k^{j,\ell}, \xi_{k+1}^i)) / \tilde{N}$.
end for
end for
end for

4 Convergence results

Consider the following assumptions.

- H1** (i) For all $k \geq 0$ and all $x \in \mathbb{R}^d$, $g_k(x) > 0$.
 (ii) $\sup_{k \geq 0} |g_k|_\infty < \infty$.

Assumption **H1** only involves the marginal likelihood g_k of the observations and does not depend on the unbiased estimation of the transition density. In the case where the observations are given as in the Section 5, this assumption holds as soon as the variance of the observation is bounded away from zero.

- H2** $\sup_{k \geq 1} |\vartheta_k|_\infty < \infty$, $\sup_{k \geq 1} |p_k|_\infty < \infty$, and $\sup_{k \geq 1} |\widehat{\omega}_k|_\infty < \infty$,
 where

$$\widehat{\omega}_0(x) = \frac{\chi(x)g_0(x)}{\eta_0(x)} \quad \text{and for } k \geq 1,$$

$$\widehat{\omega}_k(x, x'; z) = \frac{\widehat{q}_{k-1}(x, x'; z) g_k(x')}{\vartheta_k(x)p_{k-1}(x, x')}.$$

Assumption **H2** depends on the algorithm used to estimate the transition densities and on the tuning parameters of the SMC filter. The most common choice is $\vartheta_k = 1$ so that under **H1**, the only requirement is to control \widehat{q}_{k-1} and

p_{k-1} . For instance, in the case of the GPE-1, as explained in Section 3, **H2** is satisfied if ϕ is upper bounded (as for the EA1).

Lemma 2 For all $0 \leq k \leq n - 1$, the random variables $\{\widehat{\omega}_{k+1}^i \tau_{k+1}^i\}_{i=1}^N$ are independent conditionally on \mathcal{F}_k^N and

$$\mathbb{E}[\widehat{\omega}_{k+1}^1 \tau_{k+1}^1 | \mathcal{F}_k^N] = (\phi_k^N[\vartheta_{k+1}])^{-1} \phi_k^N \left[\int q_k(\cdot, x) g_{k+1}(x) \{\tau_k(\cdot) + h_{k+1}(\cdot, x)\} dx \right].$$

Proof See Appendix □

Proposition 1 Assume that **H1** and **H2** hold and that for all $1 \leq k \leq n$, $\text{osc}(h_k) < +\infty$. For all $0 \leq k \leq n$ and all $\tilde{N} \geq 1$, there exist $b_k, c_k > 0$ such that for all $N \geq 1$ and all $\varepsilon \in \mathbb{R}_+^*$,

$$\mathbb{P}(|\phi_k^N[\tau_k] - \phi_k[T_k h_k]| \geq \varepsilon) \leq b_k \exp(-c_k N \varepsilon^2).$$

Proof See Appendix □

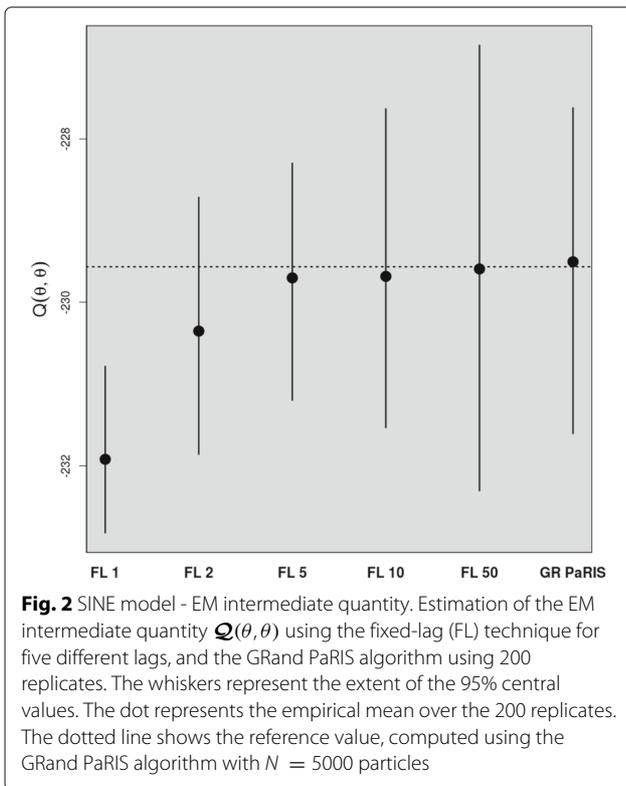
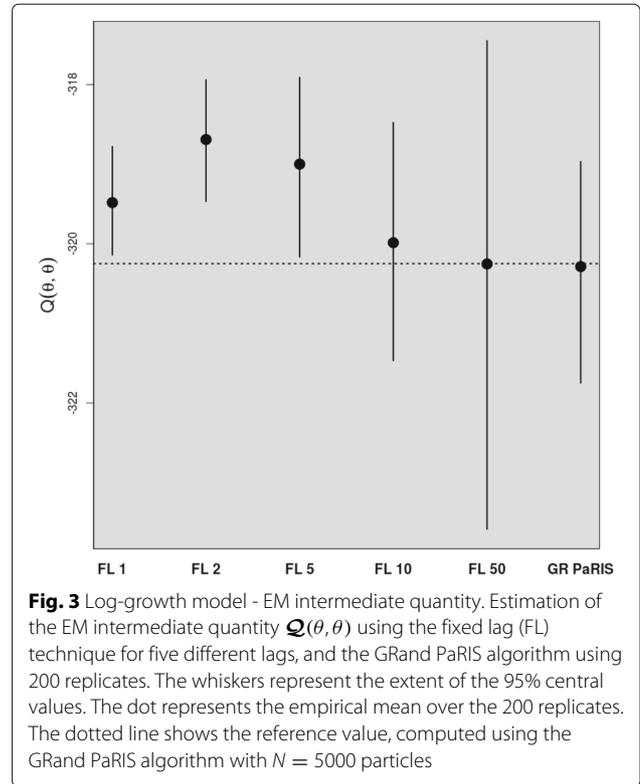
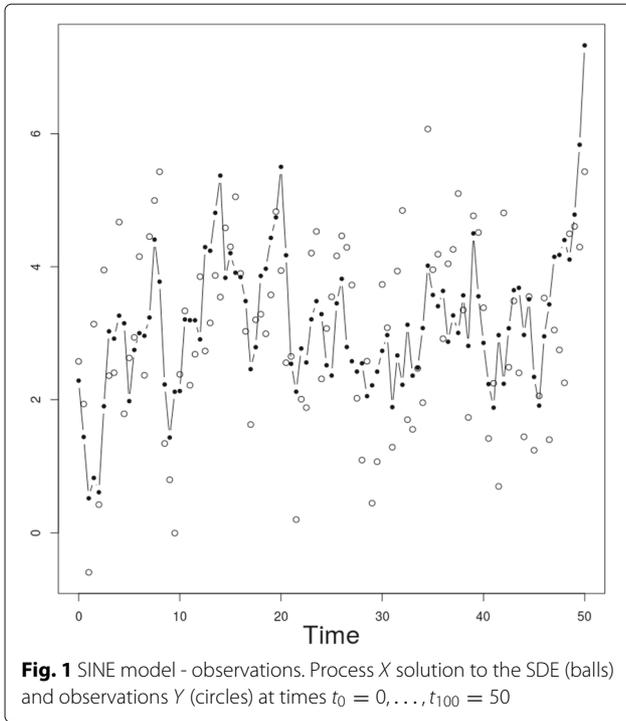
5 Numerical experiments

This section investigates the performance of the proposed algorithm with the sine and log-growth models (Fig. 1). In both cases, the proposal distribution p_k is chosen as the following approximation of the optimal filter (or the fully adapted particle filter in the terminology of [30]):

$$p_{k-1}(x_{k-1}, x_k) \propto \tilde{q}_{k-1}(x_{k-1}, x_k) g_k(x_k),$$

where $\tilde{q}_{k-1}(x_{k-1}, x_k)$ is the p.d.f. of Gaussian distribution with mean $\alpha(x_{k-1})\Delta_k$ and variance $\Delta_k I_d$, i.e., the Euler approximation of Eq. (1). As the observation model is linear and Gaussian, the proposal distribution is therefore Gaussian with explicit mean and variance.

In order to evaluate the performance of the proposed algorithm, the following strategy has been chosen. We compare the estimation of the EM intermediate quantity with the one obtained by the fixed lag method of [27], for different values of the lag (namely, 1,2,5,10,50). The particle approximation of $\mathcal{Q}(\theta, \theta)$ for each model is computed using each algorithm, see Fig. 2 for the SINE model and Fig. 3 for the log-growth model. This estimation is performed 200 times to obtain the estimates $\widehat{Q}_1, \dots, \widehat{Q}_{200}$, using $\tilde{N} = 2$ particles for PaRIS algorithm, and $M = 30$ replications for the Monte Carlo approximation \widehat{q}_k of each q_k . Moreover, the E step requires the computation of a quantity such as (2) with $h_k = \log g_k + \log q_k$. $\log q_k$ is not available explicitly and is approximated using the unbiased estimator proposed in [27, Appendix B] based on 30 independent Monte Carlo simulations. In order to obtain a reference value for our study, the intermediate quantity of the EM algorithm is also estimated 30 times using the GRand PaRIS algorithm with $N = 5000$ particles, the



reference value is then computed as the arithmetic mean of these 30 estimations, and denoted by \hat{Q}_* . Figures 2 and 3 display this estimate for an example with one simulated data set. The GRand Paris algorithm is performed using $N = 400$ particles in both cases, the fixed lag technique using $N = 1600$ so that both estimations require similar computational times, resulting a fair comparison. On a personal computer¹, for the parameters mentioned above, it takes around 25 s to perform each E step.

5.1 The SINE model

The performance of the GRand Paris algorithm are first highlighted using the SINE model, where $(X_t)_{t \geq 0}$ is supposed to be the solution to

$$dX_t = \sin(X_t - \mu) dt + dW_t, \quad X_0 = x_0. \tag{15}$$

This simple model has no explicit transition density, however, GPEs may be computed by simulating Brownian bridges. The process solution to (15) is observed regularly at times $t_0 = 0, \dots, t_{100} = 50$ through the observation process $(Y_k)_{0 \leq k \leq 100}$:

$$Y_k = X_k + \varepsilon_k,$$

where the $(\varepsilon_k)_{0 \leq k \leq 100}$ are i.i.d. $\mathcal{N}(0, \sigma_{\text{obs}}^2)$, the resulting set of model parameters is $\theta = (\mu, \sigma_{\text{obs}})$. In the example displayed in Fig. 1, we set $\mu = 0$ and $\sigma_{\text{obs}} = 1$.

In the case of the SINE model, the estimator \hat{q}_k defined by Eq. (14) satisfies both (10) and (11). The corresponding bound $\hat{\sigma}_+^k$ can be obtained using numerical optimization. If that bound is chosen, the GRand PaRIS algorithm has linear complexity in the number of particles. As an alternative, it is worth noting here that the bounds $\hat{\sigma}_+^{k,i}$, $1 \leq i \leq N$, defined by (12) can also be used. This method has a quadratic cost in the number of particles but provides the optimal bound for the algorithm of Lemma 1. This may reduce significantly the expected time before acceptance, in particular when the time step Δ_k is large. In the experiment configuration presented here, both bounds resulted in an equivalent computational time.

This same experiment was reproduced on 100 different simulated data sets. For each simulation s , the empirical absolute relative bias arb_s and the empirical absolute coefficient of variation acv_s are computed as

$$arb_s = \frac{|m(\hat{Q}^s) - \hat{Q}_*^s|}{|\hat{Q}_*^s|},$$

$$acv_s = \frac{\sigma(\hat{Q}^s)}{|m(\hat{Q}^s)|},$$

where $m(\hat{Q}^s)$ and $\sigma(\hat{Q}^s)$ are the empirical mean and standard deviation of the sequence Q_1^s, \dots, Q_{200}^s . For each estimation method, the resulting distributions of arb_1, \dots, arb_{100} and acv_1, \dots, acv_{100} are shown in Figs. 4 and 5.

The GRand PaRIS algorithm outperforms the fixed-lag methods for any value of the lag as the bias is the lowest (it is already negligible for $N = 400$) and with a lower

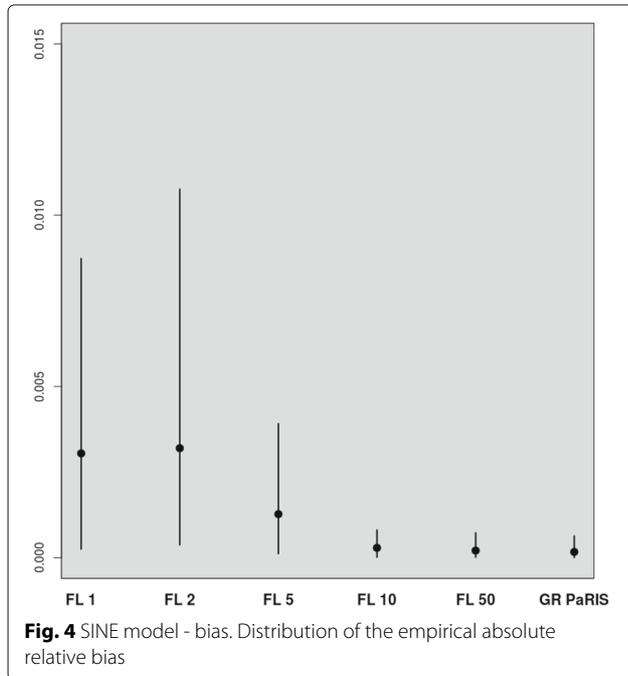


Fig. 4 SINE model - bias. Distribution of the empirical absolute relative bias

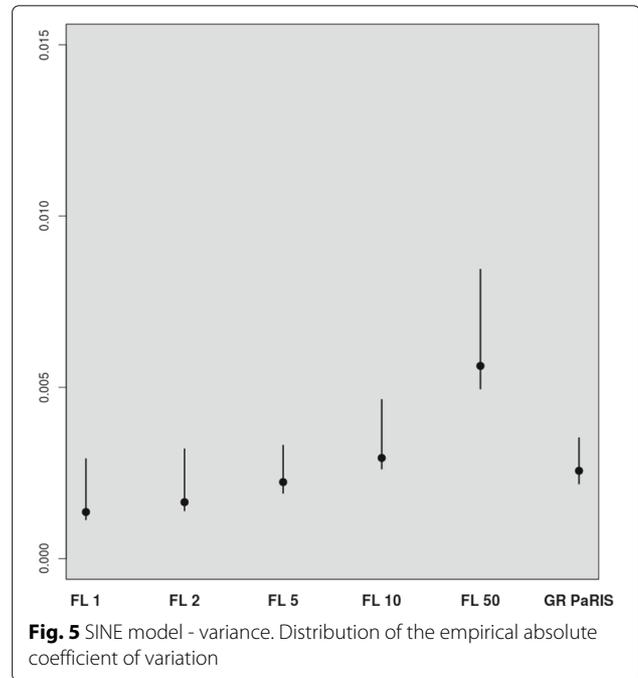


Fig. 5 SINE model - variance. Distribution of the empirical absolute coefficient of variation

variance than fixed lag estimates with negligible bias (i.e., in this case, lags larger than 10). Small lags lead to strongly biased estimates for the fixed-lag method, and unbiased estimates are at the cost of a large variance. It is worth noting here that the lag for which the bias is small is model dependent.

Generalized EM procedure The performance of our algorithm is also assessed in the case where θ and the variance σ_{obs}^2 are unknown and estimated using a generalized EM algorithm. The study is done using a data set with $n = 200$ observations simulated with $\mu = 0$ and $\sigma_{\text{obs}}^2 = 1$. The GRand PaRIS algorithm is used to perform the E step, with the same settings as before for N, \tilde{N} , and M . As there is no closed form solution to compute the M step of the EM algorithm and propose new parameter estimates, we use a generalized EM procedure: given the current estimation $\theta^{(k)} := (\mu^{(k)}, \sigma_{\text{obs}}^{(k)})$, the function $Q(\cdot, \theta^{(k)})$ is approximated for 50 new candidates $\theta_1, \dots, \theta_{50}$ chosen by the user. The new estimate is set as

$$\theta^{(k+1)} = \operatorname{argmax}_i Q(\theta_i, \theta^{(k)}).$$

This procedure has the nice property of using the same particle filter and the same retrospective sampling of Lemma 1 for all candidates, avoiding to repeat this time consuming procedure. The number of candidates and the way to choose them is problem dependent and then left to the user. In our case, we sampled candidates using Gaussian distributions around the current estimate $\theta^{(k)}$, decreasing the variance when k increases. Figures 6 and 7

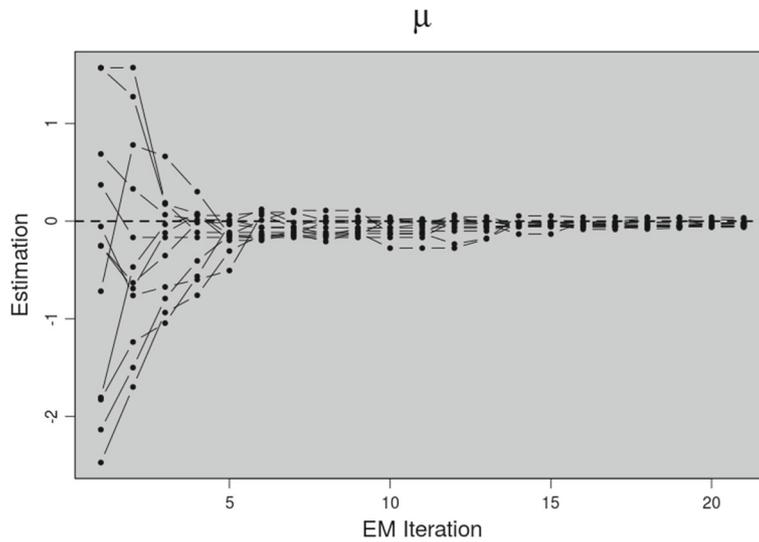


Fig. 6 SINE model - EM. Estimation of μ

illustrate the performance of the estimation for 12 different initializations of μ (resp. σ_{obs}) uniformly chosen in $]-\pi, \pi[$ (resp. in $]0, 6[$), illustrating a convergence after only a few iterations of the EM procedure.

5.2 Log-growth model

Following [6] and [28], the performance of the proposed algorithm are also illustrated with the log-growth model (Fig. 8) defined by

$$dZ_t = \kappa Z_t \left(1 - \frac{Z_t}{\gamma} \right) dt + \sigma Z_t dW_t, \quad Z_0 = z_0. \quad (16)$$

In order to use the exact algorithms of [6] and the GPE of [13], we consider (16) after the Lamperti transform, i.e., the process defined by $X_t = \eta(Z_t)$, with $\eta(z) := -\log(z)/\sigma$, which satisfies the following SDE:

$$dX_t = \left(\frac{\sigma}{2} - \frac{\kappa}{\sigma} + \frac{\kappa}{\gamma\sigma} \exp(-\sigma X_t) \right) dt + dW_t, \quad X_0 = x_0 = \eta(z_0). \quad (17)$$

In this case, the conditions of the exact Algorithm 2 defined in [6] are satisfied, as for any $m \in \mathbb{R}$ there exists

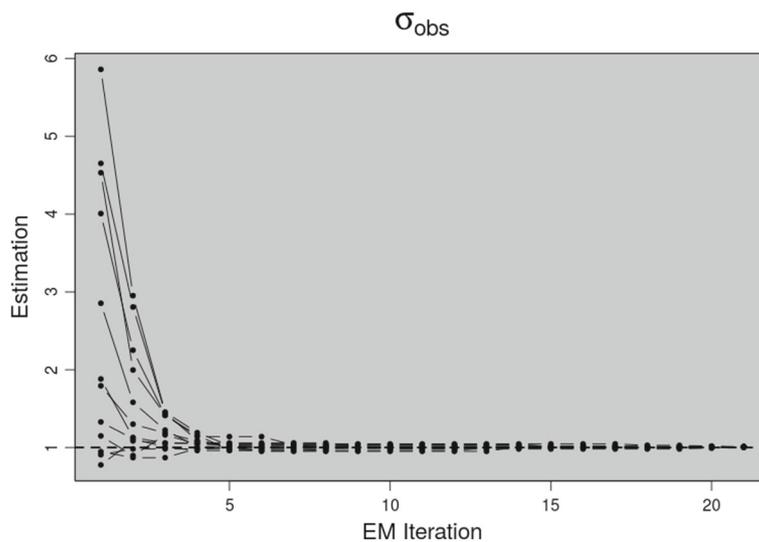
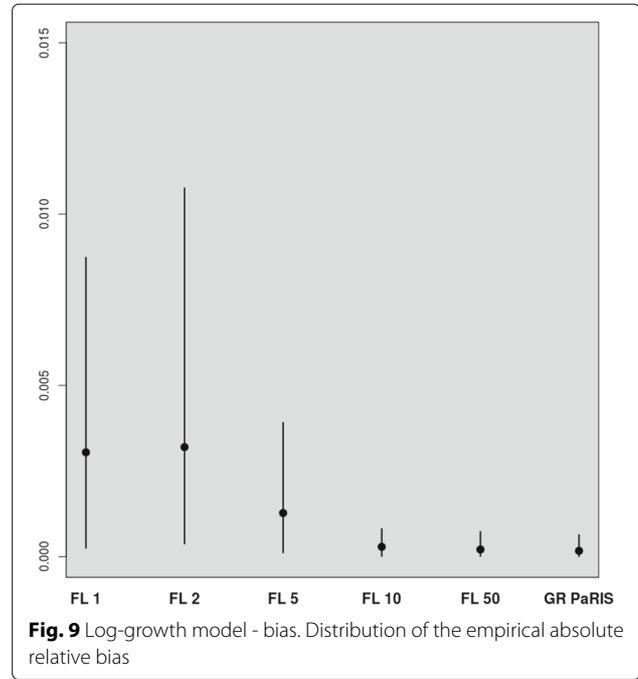
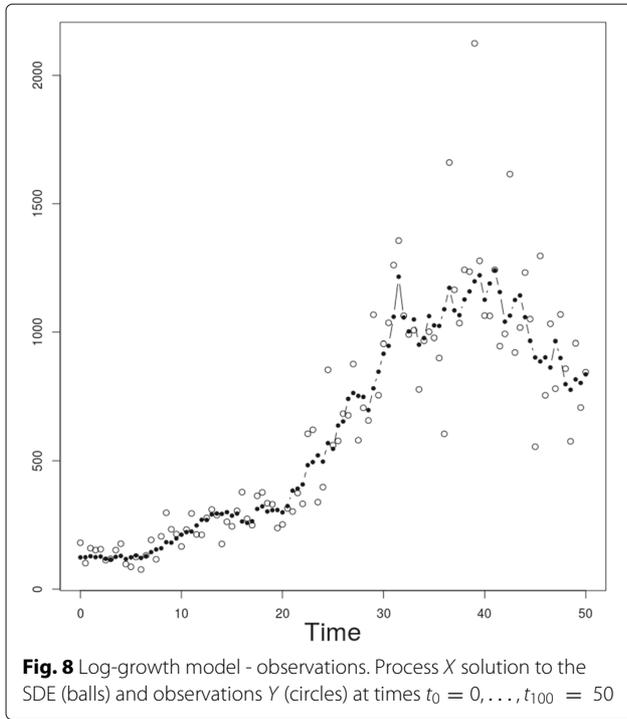


Fig. 7 SINE model - EM. Estimation of σ_{obs}



U_m such that for all $x \geq m$, $\psi(x) := \alpha^2(x) + \alpha'(x) \leq U_m$. Moreover, ψ is lower bounded uniformly by L . Then, GPE estimators may be computed by simulating the minimum of a Brownian bridge, and simulating Bessel bridges conditionally to this minimum, as proposed by [6].

The process solution to (17) is observed regularly at times $t_0 = 0, \dots, t_{50} = 100$ through the observation process $(Y_k)_{0 \leq k \leq 50}$ defined as

$$Y_k = X_k + \varepsilon_k,$$

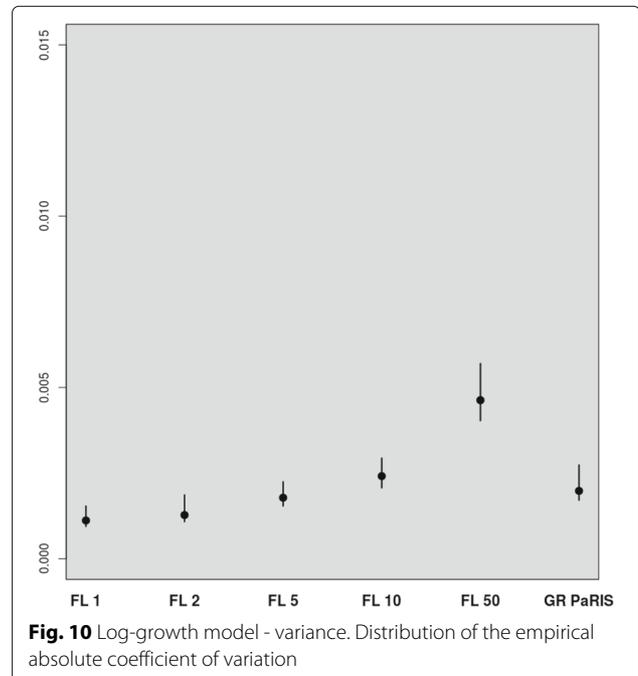
where the $(\varepsilon_k)_{0 \leq k \leq 50}$ are i.i.d. $\mathcal{N}(0, \sigma_{obs}^2)$. The parameters are given by

$$\theta = (\kappa = 0.1, \sigma = 0.1, \gamma = 1000, \sigma_{obs} = 2).$$

In the case of the log-growth model, the estimator $\hat{q}_k(\cdot)$ defined by Eq. (14) satisfies (11), leading to a GRand PaRIS algorithm with linear complexity in the number of particles. However, the remarks about the bound $\hat{\sigma}_+^k$ made for the SINE model above still hold in this case. The intermediate quantity of the EM algorithm is evaluated as for the SINE model, see Figs. 3, 9, and 10.

The results for the fixed-lag technique are similar to the ones presented in [27, Figure 1] using the same model. For small lags, the variance of the estimates is small, but the estimation is highly biased. The bias rapidly decreases as the lag increases, together with a great increase of variance. Again, the GRand PaRIS algorithm outperforms the fixed lag smoother as it shows a similar (vanishing) bias as the fixed lag for the largest lag and a smaller variance than the fixed lags estimates with negligible bias.

Note that in this case, the Lamperti transform to obtain a diffusion with a unitary diffusion term depends on σ . The process $(X_t)_{t \geq 0}$ is a function of σ and is not directly observed if σ is unknown, which prevents a direct use of an EM algorithm to estimate σ . Following [6, Section 8.2], this may be overcome with a two-step transformation of the process $(Z_t)_{t \geq 0}$.



6 Conclusions

This paper presents a new online SMC smoother for partially observed differential equations. This algorithm relies on an acceptance-rejection procedure inspired from the recent PaRIS algorithm. The main result of the article for practical applications is that the mechanism of this procedure remains valid when the transition density is approximated by an unbiased positive estimator. The proposed procedure therefore extends the PaRIS algorithm to HMMs whose transition density is unknown and can be unbiasedly approximated. The GRand PaRIS algorithm outperforms the existing fixed lag smoother for POD processes of [27], as it does not introduce any intrinsic and non-vanishing bias. In addition, numerical simulations highlight a better variance using data from two different models. It can be implemented for the class of models for which exact algorithms of [6] are valid, with a linear complexity in N in the best cases, or at worse in N^2 .

Endnote

¹ (i7-6600U CPU @ 2.60GHz)

Appendix

Proofs

Proof of Lemma 1 Let τ be the first time draws are accepted in the accept-reject mechanism. For all $\ell \geq 1$, write

$$\mathcal{A}_\ell^k = \left\{ U_\ell < \widehat{q}_k \left(\xi_k^\ell, \xi_{k+1}^i, \zeta_k^\ell \right) / \widehat{\sigma}_+^k \right\}.$$

Let h be a function defined on $\{1, \dots, N\}$,

$$\begin{aligned} & \mathbb{E} \left[h \left(J_k^{ij} \right) \middle| \mathcal{G}_{k+1}^N \right] \\ &= \sum_{m \geq 1} \mathbb{E} \left[h(J_m) \mathbf{1}_{\tau=m} \middle| \mathcal{G}_{k+1}^N \right], \\ &= \sum_{m \geq 1} \left(\prod_{\ell=1}^{m-1} \mathbb{E} \left[\mathbf{1}_{(\mathcal{A}_\ell^k)^c} \middle| \mathcal{G}_{k+1}^N \right] \right) \mathbb{E} \left[h(J_m) \mathbf{1}_{\mathcal{A}_m^k} \middle| \mathcal{G}_{k+1}^N \right], \\ &= \sum_{m \geq 1} \left(\prod_{\ell=1}^{m-1} \mathbb{E} \left[1 - \frac{\widehat{q}_k \left(\xi_k^\ell, \xi_{k+1}^i, \zeta_k^\ell \right)}{\widehat{\sigma}_+^k} \middle| \mathcal{G}_{k+1}^N \right] \right) \\ & \quad \times \mathbb{E} \left[h(J_m) \frac{\widehat{q}_k \left(\xi_k^m, \xi_{k+1}^i, \zeta_k^m \right)}{\widehat{\sigma}_+^k} \middle| \mathcal{G}_{k+1}^N \right], \\ &= \sum_{m \geq 1} \left(\mathbb{E} \left[1 - \frac{q_k \left(\xi_k^1, \xi_{k+1}^1 \right)}{\widehat{\sigma}_+^k} \middle| \mathcal{G}_{k+1}^N \right] \right)^{m-1} \\ & \quad \mathbb{E} \left[h(J_1) \frac{q_k \left(\xi_k^1, \xi_{k+1}^1 \right)}{\widehat{\sigma}_+^k} \middle| \mathcal{G}_{k+1}^N \right], \\ &= \mathbb{E} \left[h(J_1) q_k \left(\xi_k^1, \xi_{k+1}^1 \right) \middle| \mathcal{G}_{k+1}^N \right] / \mathbb{E} \left[q_k \left(\xi_k^1, \xi_{k+1}^1 \right) \middle| \mathcal{G}_{k+1}^N \right], \\ &= \sum_{\ell=1}^N \frac{h(\ell) \omega_{k-1}^\ell q_k \left(\xi_k^\ell, \xi_{k+1}^i \right)}{\sum_{m=1}^N \omega_{k-1}^m q_k \left(\xi_k^m, \xi_{k+1}^i \right)}, \\ &= \sum_{\ell=1}^N \Lambda_{k-1}^N(i, \ell) h(\ell), \end{aligned}$$

which concludes the proof.

Proof of Lemma 2 The independence is ensured by the mechanism of SMC methods. By (9),

$$\begin{aligned} & \mathbb{E} \left[\widehat{\omega}_{k+1}^i \tau_{k+1}^i \middle| \mathcal{F}_k^N \right] = \\ & \mathbb{E} \left[\frac{\widehat{q}_k \left(\xi_k^{i_{k+1}}, \xi_{k+1}^i; \zeta_k \right) g_{k+1} \left(\xi_{k+1}^i \right)}{\vartheta_{k+1} \left(\xi_k^{i_{k+1}} \right) p_k \left(\xi_k^{i_{k+1}}, \xi_{k+1}^i \right)} \tau_{k+1}^i \middle| \mathcal{F}_k^N \right]. \end{aligned}$$

Note that by Lemma 1,

$$\begin{aligned} & \mathbb{E} \left[\tau_{k+1}^i \middle| \mathcal{G}_{k+1}^N \right] = \sum_{\ell=1}^N \frac{\omega_k^\ell q_k \left(\xi_k^\ell, \xi_{k+1}^i \right) \left(\tau_k^\ell + h_k \left(\xi_k^\ell, \xi_{k+1}^i \right) \right)}{\sum_{\ell'=1}^N \omega_k^{\ell'} q_k \left(\xi_k^{\ell'}, \xi_{k+1}^i \right)}, \\ & \mathbb{E} \left[\widehat{q}_k \left(\xi_k^{i_{k+1}}, \xi_{k+1}^i; \zeta_k \right) \middle| \mathcal{G}_{k+1}^N \right] = q_k \left(\xi_k^{i_{k+1}}, \xi_{k+1}^i \right). \end{aligned}$$

Since τ_{k+1}^i and ζ_k are independent conditionally to \mathcal{G}_{k+1}^N :

$$\begin{aligned} & \mathbb{E} \left[\tau_{k+1}^i \widehat{q}_k \left(\xi_k^{i_{k+1}}, \xi_{k+1}^i; \zeta_k \right) \middle| \mathcal{G}_{k+1}^N \right] \\ &= q_k \left(\xi_k^{i_{k+1}}, \xi_{k+1}^i \right) \sum_{\ell=1}^N \frac{\omega_k^\ell q_k \left(\xi_k^\ell, \xi_{k+1}^i \right) \left(\tau_k^\ell + h_k \left(\xi_k^\ell, \xi_{k+1}^i \right) \right)}{\sum_{\ell'=1}^N \omega_k^{\ell'} q_k \left(\xi_k^{\ell'}, \xi_{k+1}^i \right)}. \end{aligned}$$

Moreover, conditionally to \mathcal{F}_k^N , the probability density function of $\left(\xi_{k+1}^i, I_{k+1}^i \right)$ is given by

$$(x, j) \mapsto \frac{\omega_k^j \vartheta_{k+1} \left(\xi_k^j \right) p_k \left(\xi_k^j, x \right)}{\Omega_k^N \phi_k^N \left[\vartheta_{k+1} \right]}.$$

Therefore, this yields

$$\begin{aligned} & \mathbb{E} \left[\widehat{\omega}_{k+1}^i \tau_{k+1}^i \middle| \mathcal{F}_k^N \right] \\ &= \left(\phi_k^N \left[\vartheta_{k+1} \right] \right)^{-1} \sum_{j=1}^N \frac{\omega_k^j}{\Omega_k} \int \vartheta_{k+1} \left(\xi_k^j \right) \frac{q_k \left(\xi_k^j, x \right) g_{k+1}(x)}{\vartheta_{k+1} \left(\xi_k^j \right) p_k \left(\xi_k^j, x \right)} \\ & \quad \times \sum_{\ell=1}^N \frac{\omega_k^\ell q_k \left(\xi_k^\ell, x \right) \left(\tau_k^\ell + h_k \left(\xi_k^\ell, x \right) \right)}{\sum_{\ell'=1}^N \omega_k^{\ell'} q_k \left(\xi_k^{\ell'}, x \right)} p_k \left(\xi_k^j, x \right) dx, \\ &= \left(\phi_k^N \left[\vartheta_{k+1} \right] \right)^{-1} \\ & \quad \times \sum_{\ell=1}^N \frac{\omega_k^\ell}{\Omega_k} \left[\int \frac{\sum_{j=1}^N \omega_k^j q_k \left(\xi_k^j, x \right)}{\sum_{\ell'=1}^N \omega_k^{\ell'} q_k \left(\xi_k^{\ell'}, x \right)} g_{k+1}(x) q_k \left(\xi_k^\ell, x \right) \right. \\ & \quad \left. \left(\tau_k^\ell + h_k \left(\xi_k^\ell, x \right) \right) dx \right] \\ &= \left(\phi_k^N \left[\vartheta_{k+1} \right] \right)^{-1} \phi_k^N \left[\int q_k(\cdot, x) g_{k+1}(x) \{ \tau_k(\cdot) + h_k(\cdot, x) \} dx \right], \end{aligned}$$

which concludes the proof. \square

Proof of Proposition 1 The results is proved by induction. At time $k = 0$, the result holds using that for all $1 \leq i \leq N$, $\rho_0^i = 0$ and the convention $T_0[h_0] = 0$. In

\square

addition, ϕ_0^N is a standard importance sampler estimator of ϕ_0 with $\widehat{\omega}_0^i \leq |\widehat{\omega}_0|_\infty$ so that for any bounded function h on X ,

$$\mathbb{P}(|\phi_0^N[h] - \phi_0[h]| \geq \varepsilon) \leq b_0 \exp(-c_0 N \varepsilon^2).$$

Assume the results holds for $k \geq 0$ and that $\vartheta_{k+1} = 1$ for simplicity. Write

$$\phi_{k+1}^N[\tau_{k+1}] - \phi_{k+1}[T_{k+1}[h_{k+1}]] = a_N/b_N,$$

where $a_N = N^{-1} \sum_{i=1}^N \widehat{\omega}_{k+1}^i (\tau_{k+1}^i - \phi_{k+1}[T_{k+1}[h_{k+1}]])$ and $b_N = N^{-1} \sum_{i=1}^N \widehat{\omega}_{k+1}^i$. By Lemma 2, the random variables $\{\widehat{\omega}_{k+1}^i \tau_{k+1}^i\}_{i=1}^N$ are independent conditionally on \mathcal{F}_k^N and by H2,

$$|\widehat{\omega}_{k+1}^i (\tau_{k+1}^i - \phi_{k+1}[T_{k+1}[h_{k+1}]])| \leq 2|\widehat{\omega}_{k+1}|_\infty |H_{k+1}|_\infty.$$

Therefore, by Hoeffding inequality,

$$\mathbb{P}(|a_N - \mathbb{E}[a_N | \mathcal{F}_k^N]| \geq \varepsilon) = \mathbb{E}[\mathbb{P}(|a_N - \mathbb{E}[a_N | \mathcal{F}_k^N]| \geq \varepsilon | \mathcal{F}_k^N)] \leq 2 \exp(-c_k N \varepsilon^2).$$

On the other hand,

$$\mathbb{E}[a_N | \mathcal{F}_k^N] = \phi_k^N[\Upsilon_k],$$

where

$$\Upsilon_k(x_k) = \int q_k(\cdot, x) g_{k+1}(x) (\tau_k(x_k) + h_{k+1}(x_k, x) - \phi_{k+1}[T_{k+1}[h_{k+1}]]) dx.$$

By [28, Lemma 11], $\phi_k[\Upsilon_k] = 0$ which implies by the induction assumption that

$$\mathbb{P}(|\mathbb{E}[a_N | \mathcal{F}_k^N]| \geq \varepsilon) \leq b_k \exp(-c_k N \varepsilon^2).$$

Then,

$$\mathbb{P}(|a_N| \geq \varepsilon) \leq b_k \exp(-c_k N \varepsilon^2).$$

Similarly, as $b_N \leq |\widehat{\omega}_k|_\infty$, by Hoeffding inequality,

$$\mathbb{P}(|b_N - \mathbb{E}[b_N | \mathcal{F}_k^N]| \geq \varepsilon) = \mathbb{E}[\mathbb{P}(|b_N - \mathbb{E}[b_N | \mathcal{F}_k^N]| \geq \varepsilon | \mathcal{F}_k^N)] \leq 2 \exp(-c_k N \varepsilon^2).$$

Note that

$$\mathbb{E}[b_N | \mathcal{F}_k^N] = \phi_k^N \left[\int q_k(\cdot, x) g_{k+1}(x) dx \right].$$

By the induction assumption,

$$\mathbb{P} \left(\left| \mathbb{E}[b_N | \mathcal{F}_k^N] - \phi_k \left[\int q_k(\cdot, x) g_{k+1}(x) dx \right] \right| \geq \varepsilon \right) \leq b_k \exp(-c_k N \varepsilon^2).$$

The proof is completed using Lemma 3. □

Lemma 3 Assume that a_N , b_N , and b are random variables defined on the same probability space such that there exist positive constants β , B , C , and M satisfying

- (i) $|a_N/b_N| \leq M$, \mathbb{P} -a.s. and $b \geq \beta$, \mathbb{P} -a.s.,
- (ii) For all $\epsilon > 0$ and all $N \geq 1$, $\mathbb{P}[|b_N - b| > \epsilon] \leq B \exp(-CN\epsilon^2)$,
- (iii) For all $\epsilon > 0$ and all $N \geq 1$, $\mathbb{P}[|a_N| > \epsilon] \leq B \exp(-CN(\epsilon/M)^2)$.

Then,

$$\mathbb{P} \left\{ \left| \frac{a_N}{b_N} \right| > \epsilon \right\} \leq B \exp \left(-CN \left(\frac{\epsilon \beta}{2M} \right)^2 \right).$$

Proof See [10]. □

Funding

This work has been developed during a 1-year postdoc funded by Paris-Saclay Center for Data Science.

Authors' contributions

All the authors have contributed to the conception of the algorithms, the analysis of the proposed estimator, and to the redaction of the manuscript. PG provided the simulations displayed in the final version. All authors read and approved the final manuscript.

Competing interests

The authors declare that they have no competing interests.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Author details

¹AgroParisTech, UMR MIA 518, F-75231, Paris, France. ²Laboratoire de Mathématiques d'Orsay, Univ. Paris-Sud, CNRS, Université Paris-Saclay, 91405 Orsay, France.

Received: 5 March 2017 Accepted: 11 January 2018

Published online: 02 February 2018

References

1. Y Ait-Sahalia, Transition densities for interest rate and other nonlinear diffusions. *J. Financ.* **54**, 1361–1395 (1999)
2. Y Ait-Sahalia, Maximum likelihood estimation of discretely sampled diffusions: a closed-form approximation approach. *Econometrica.* **70**, 223–262 (2002)
3. Y Ait-Sahalia, Closed-form likelihood expansions for multivariate diffusions. *Ann. Stat.* **36**, 906–937 (2008)
4. A Beskos, O Papaspiliopoulos, GO Roberts, Retrospective exact simulation of diffusion sample paths with applications. *Bernoulli.* **12**(6), 1077:1098 (2006)
5. A Beskos, O Papaspiliopoulos, GO Roberts, A factorisation of diffusion measure and finite sample path constructions. *Methodol. Comput. Appl. Probab.* **10**(1), 85–104 (2008)
6. A Beskos, O Papaspiliopoulos, GO Roberts, P Fearnhead, Exact and computationally efficient likelihood-based estimation for discretely observed diffusion processes (with discussion). *J. Roy. Statist. Soc. Ser. B.* **68**(3), 333–382 (2006)
7. O Cappé, E Moulines, T Rydén, *Inference in hidden Markov models.* (Springer-Verlag, New York, 2005)
8. P Del Moral, J Jacod, P Protter, The Monte Carlo method for filtering with discrete-time observations. *Probab. Theory Relat. Fields.* **120**, 346–368 (2001)
9. AP Dempster, NM Laird, DB Rubin, Maximum likelihood from incomplete data via the EM algorithm. *J. Roy. Statist. Soc. B.* **39**(1), 1–38 (1977). (with discussion)
10. R Douc, A Garivier, E Moulines, J Olsson, Sequential Monte Carlo smoothing for general state space hidden Markov models. *Ann. Appl. Probab.* **21**(6), 2109–2145 (2011)

11. A Doucet, S Godsill, C Andrieu, On sequential Monte-Carlo sampling methods for Bayesian filtering. *Stat. Comput.* **10**, 197–208 (2000)
12. A Doucet, S Godsill, C Andrieu, On sequential monte-carlo sampling methods for bayesian filtering. *Stat. Comput.* **10**, 197–208 (2000)
13. P Fearnhead, O Papaspiliopoulos, GO Roberts, Particle filters for partially observed diffusions. *J. Roy. Statist. Soc. Ser. B.* **70**(4), 755–777 (2008)
14. P Fearnhead, K Latuszynski, GO Roberts, G Sermaidis, Continuous-time importance sampling: Monte Carlo methods which avoid time-discretisation error (2017). Technical report
15. P Gloaguen, M-P Étienne, S Le Corff, Stochastic differential equation based on a multimodal potential to model movement data in ecology. To appear in the *Journal of the Royal Statistical Society: Series C.* <http://onlinelibrary.wiley.com/doi/10.1111/rssc.12251/abstract>
16. SJ Godsill, A Doucet, M West, Monte Carlo smoothing for non-linear time series. *J. Am. Stat. Assoc.* **50**, 438–449 (2004)
17. N Gordon, D Salmond, AF Smith, Novel approach to nonlinear/non-Gaussian bayesian state estimation. *IEE Proc. F. Radar Sig. Process.* **140**, 107–113 (1993)
18. M Hürzeler, HR Künsch, Monte Carlo approximations for general state-space models. *J. Comput. Graph. Stat.* **7**, 175–193 (1998)
19. N Kantas, A Doucet, SS Singh, J Maciejowski, N Chopin, On particle methods for parameter estimation in state-space models. *Stat. Sci.* **30**(3), 328–351 (2015)
20. M Kessler, Estimation of an ergodic diffusion from discrete observations. *Scand. J. Stat.* **24**(2), 211–229 (1997)
21. M Kessler, A Lindner, M Sorensen, *Statistical methods for stochastic differential equations.* (CRC Press, Boca Raton, 2012)
22. G Kitagawa, Monte-Carlo filter and smoother for non-Gaussian nonlinear state space models. *J. Comput. Graph. Stat.* **1**, 1–25 (1996)
23. S Le Corff, G Fort, Convergence of a particle-based approximation of the block online Expectation Maximization algorithm. *ACM Trans. Model. Comput. Simul.* **23**(1), 2 (2013)
24. S Le Corff, G Fort, Online expectation maximization based algorithms for inference in hidden Markov models. *Electron. J. Stat.* **7**, 763–792 (2013)
25. C Li, Maximum-likelihood estimation for diffusion processes via closed-form density expansions. *Ann. Stat.* **41**(3), 1350–1380 (2013)
26. J Olsson, O Cappe, R Douc, E Moulines, Sequential monte carlo smoothing with application to parameter estimation in nonlinear state space models. *Bernoulli.* **14**(1), 155–179 (2008)
27. J Olsson, J Strojby, Particle-based likelihood inference in partially observed diffusion processes using generalised Poisson estimators. *Electron. J. Stat.* **5**, 1090–1122 (2011)
28. J Olsson, J Westerborn, Efficient particle-based online smoothing in general hidden Markov models: the PaRIS algorithm. *Bernoulli.* **3**, 1951–1996 (2017)
29. T Ozaki, A bridge between nonlinear time series models and nonlinear stochastic dynamical systems: a local linearization approach. *Stat. Sin.* **2**, 1130–1135 (1992)
30. MK Pitt, N Shephard, Filtering via simulation: Auxiliary particle filters. *J. Am. Stat. Assoc.* **94**(446), 590–599 (1999)
31. G Poyiadjis, A Doucet, SS Singh, Particle approximations of the score and observed information matrix in state space models with application to parameter estimation. *Biometrika.* **98**, 65–80 (2011)
32. I Shoji, T Ozaki, Estimation for nonlinear stochastic differential equations by a local linearization method 1. *Stoch. Anal. Appl.* **16**(4), 733–752 (1998)
33. M Uchida, N Yoshida, Adaptive estimation of an ergodic diffusion process based on sampled data. *Stoch. Process. Appl.* **122**(8), 2885–2924 (2012)
34. W Wagner, Unbiased Monte Carlo estimators for functionals of weak solutions of stochastic differential equations. *Stochast. Stochast. Rep.* **28**, 1–20 (1989)

Submit your manuscript to a SpringerOpen[®] journal and benefit from:

- Convenient online submission
- Rigorous peer review
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

Submit your next manuscript at ► springeropen.com
