[84]

P. Fearnhead, O. Papaspiliopoulos, G.O. Roberts and A.M. Stuart,

Random-weight particle filtering of continuous time stochastic processes.

Journal of the Royal Statistical Society B. 72(4)

(2010) 497-512.

Journal of the Royal Statistical Society
SERIES B Statistical Methodology

J. R. Statist. Soc. B (2010)
72, *Part* 4, *pp.* 497–512

Random-weight particle filtering of continuous time processes

Paul Fearnhead,

Lancaster University, UK

Omiros Papaspiliopoulos

Universitat Pompeu Fabra, Barcelona, Spain

and Gareth O. Roberts and Andrew Stuart

University of Warwick, Coventry, UK

[Received January 2008. Final revision January 2010]

Summary. It is possible to implement importance sampling, and particle filter algorithms, where the importance sampling weight is random. Such random-weight algorithms have been shown to be efficient for inference for a class of diffusion models, as they enable inference without any (time discretization) approximation of the underlying diffusion model. One difficulty of implementing such random-weight algorithms is the requirement to have weights that are positive with probability 1. We show how Wald's identity for martingales can be used to ensure positive weights. We apply this idea to analysis of diffusion models from high frequency data. For a class of diffusion models we show how to implement a particle filter, which uses all the information in the data, but whose computational cost is independent of the frequency of the data. We use the Wald identity to implement a random-weight particle filter for these models which avoids time discretization error.

Keywords: Diffusions; Exact simulation; Gaussian process; Integrated processes; Negative importance weights; Poisson estimator; Sequential Monte Carlo methods

1. Introduction

The paper develops novel methods for optimal filtering of partially observed stochastic processes. There are two contributions which are combined to yield an efficient algorithm for estimating an unobserved signal when the dynamics of the signal and the observed process are described by a system of stochastic differential equations (SDEs).

The first contribution relates to so-called *random-weight importance sampling* (IS) methods. These methods are based on the simple principle that valid use of IS for stochastic simulation requires only an *unbiased estimator* of the likelihood ratio between the target and proposal distribution, rather than the ratio itself. For brevity we call such estimators *random weights*. Note, for instance, that the validity of rejection sampling follows directly from this observation, since the 0–1 weights (accept–reject) that are associated with the samples proposed are random weights. The rejection control algorithm (Liu, 2001) is similarly justified.

Address for correspondence: Omiros Papaspiliopoulos, Department of Economics, Universitat Pompeu Fabra, Ramon Trias Fargas 25–27, 08005 Barcelona, Spain.
E-mail: omiros.papaspiliopoulos@upf.edu

 1369–7412/10/72497

Of particular relevance to this paper is the application of this idea when the likelihood ratio is intractable. Intractable likelihood ratios arise naturally in inference and simulation of continuous time processes. The case of SDEs is discussed in detail in this paper; see Barndorff-Nielsen and Shephard (2001) for models that are based on Lévy processes. Additionally, it has recently emerged that the appropriate use of unbiased estimators of likelihood ratios within Markov chain Monte Carlo sampling ensures the desired limiting distribution; see for example Andrieu and Roberts (2009), Andrieu *et al.* (2010) and Beskos *et al.* (2006). Random weights have also been used in the context of option pricing; see Glasserman and Staum (2001).

Powerful techniques have been developed for obtaining random weights even in complex situations. These techniques generate unbiased estimates of non-linear functionals of a quantity under the assumption that the quantity itself can be unbiasedly estimated, e.g. unbiasedly estimating $\exp(\Phi)$ given unbiased estimates of Φ. The *Poisson estimator* and its generalizations that are reviewed in Section 3.2 provide one such technique; see Papaspiliopoulos (2009) for an overview of this methodology.

In the context of sequential IS the use of random weights was first proposed in Rousset and Doucet (2006) and Fearnhead *et al.* (2008). Fearnhead *et al.* (2008) introduced the so-called *random-weight particle filter* (RWPF) and demonstrated that substituting intractable weights with *positive* unbiased estimators is equivalent to a data augmentation approach. In sequential IS the weights are involved in resampling steps; therefore for efficient implementation it is required that they are positive and can be normalized to yield resampling probabilities. Whereas obtaining random weights might be easy by using the Poisson estimator, controlling their sign is difficult.

We propose to use Wald's identity for martingales to generate unbiased estimators which are guaranteed to be positive. The estimators are defined up to an unknown constant of proportionality, which is removed when IS weights are normalized to sum to 1. This technique can be used within a particle filter to yield what we call the Wald random-weight particle filter (WRWPF). We show that under moment conditions on the random weights this modification increases the computational complexity of the filter by only a logarithmic factor.

Secondly we develop a new approach for optimal filtering of Markov models from high frequency observations. The approach is particularly suited to the continuous time filtering problem (Del Moral and Miclo, 2000), where the observed process and the signal evolve according to a system of SDEs. Hence, for clarity Section 3 develops the new approach in this context, and we defer discussion about extensions until Section 5.

Generally, filtering for SDEs is impeded by the fact that transition densities are unavailable, except for very special cases such as linear SDEs. This intractability renders even the application of Monte Carlo methods (i.e particle filters) non-trivial. In contrast, for a certain class of SDEs (see Section 3.1) unbiased estimators of transition densities can be obtained; thus the RWPF can be applied. This approach was developed in Fearnhead *et al.* (2008) to provide *exact* (i.e. model-approximation-free) particle approximations to the filtering (and smoothing) distributions.

Here we show how to obtain particle approximations to the filtering distributions at a set of auxiliary filtering times which is a subset of the observation times. This extension is particularly important when we have high frequency observed data, where filtering at every observation time comes at a large computational cost due to the vast amount of data.

Subsampling the observed data can reduce the computational cost but it throws away information. Our new filter applies to the class of SDEs where unbiased estimators of the transition density are available and is based on the unbiased estimation of the *conditional transition density* of the signal between the filtering times given any observed data on that interval. This approach

still uses all the information within the data, yet the resulting computational cost of calculating weights for each particle does not increase with the frequency of the data. To be able to implement the resulting filter we incorporate our unbiased estimates of the conditional transitions density within a WRWPF.

The outline of the paper is as follows. Section 2 introduces the Wald random-weight IS as a generic simulation method and establishes its validity in theorem 1. Section 3 introduces the new approach to filtering partially observed diffusion processes and develops a WRWPF. Section 4 applies the new WRWPF to high frequency data from models where the signal evolves at a much slower rate than the observation process and gives comparisons with alternative methods. Section 5 closes with a discussion about the scope and extensions of the methodology proposed.

2. Wald random-weight importance sampling

We are interested in sampling from a probability distribution $\mathbb{Q}$ by using proposals from a tractable distribution $\mathbb{W}$ that is defined on the same space. Let $w(x) = \mathrm{d}\mathbb{Q}/\mathrm{d}\mathbb{W}(x)$ be the corresponding Radon–Nikodym derivative (i.e. the likelihood ratio). Then, the *weighted sample* $(x^{(j)}, w^{(j)})$, $j = 1, \ldots, N$, with $w^{(j)} = w(x^{(j)})$, and $x^{(j)} \sim \mathbb{W}$, constitutes an IS approximation of $\mathbb{Q}$, and expectations $\mathbb{E}_{\mathbb{Q}}(g)$ can be consistently estimated (as $N \to \infty$) by

$$\sum_{j=1}^{N} g(x^{(j)}) \frac{w^{(j)}}{\sum_{l} w^{(l)}}. \tag{1}$$

Note that renormalization by the sum of the weights means that we need only known w up to a normalizing constant. Since the weighted sample can be used to recover arbitrary expectations, it is called a *particle approximation*. We interpret this approximation as approximating $\mathbb{Q}$ by a distribution with discrete support points, the particles $x^{(j)}$, and the probability associated with each support point is proportional to the weight $w^{(j)}$.

However, in several applications (see Section 3 and the discussion in Section 1) $w(x)$ is not explicitly available. A common situation is that $w(x) = f(x)\, r(x)$, where f is explicitly known but r is only expressed as an expectation which cannot be computed analytically. However, powerful machinery is now available (see Section 3.2) which in many cases provides an unbiased estimator of $w(x)$, say $\mathcal{W}(x)$, for any x. $\mathcal{W}$ is constructed by simulating auxiliary variables. Thus, for a given x we can obtain a sequence of estimators $\mathcal{W}_k(x)$, $k = 1, 2, \ldots$, which are conditionally independent. An important observation is that the *random weighted sample* $(x^{(j)}, \mathcal{W}^{(j)})$ is also a valid IS approximation of $\mathbb{Q}$ and expression (1) applies with w replaced by $\mathcal{W}$. This observation is a direct consequence of iterated conditional expectation. Therefore, IS can be cast in these more general terms.

However, the interpretation of IS as a distributional (particle) approximation collapses in the generalized framework when the estimators are not positive almost surely. This interpretation is fundamental when the particle approximation replaces $\mathbb{Q}$ in probability calculations, such as the application of Bayes theorem. The collection of techniques that is known as sequential Monte Carlo (SMC) sampling, a special case of which are particle filters (PFs), relies explicitly on this interpretation of IS; see Section 3 for details on PFs.

Thus, it is crucial to have a generic methodology which produces *positive random weights*, to allow the generalized IS framework to be routinely applied in SMC sampling. One simple way to achieve this is to truncate all negative weights, $\mathcal{W}^{\mathrm{trunc}} = \max\{\mathcal{W}, 0\}$. This obviously distorts the properly weighted principle by introducing a bias, but it has a desirable mean-square error property:

$$\mathbb{E}(\mathcal{W}^{\text{trunc}} - w)^2 \leqslant \mathbb{E}(\mathcal{W} - w)^2.$$

The bias will typically decrease with increasing computational cost per estimator. See Section 3.2 for further discussion in the context of the Poisson estimator. However, in practice, it is difficult to quantify the effect of the bias on the quality of the particle approximation.

We introduce a technique which, rather than setting negative weights to 0, uses extra simulation to make them positive. This is done while ensuring that the weights remain unbiased (up to a common constant of proportionality) by using the following result.

Theorem 1. Consider an infinite array of independent random variables $\mathcal{W}_k^{(j)}$ for $j = 1, \ldots, N$ and $k = 1, 2, \ldots$ each with finite fourth moment and first moment w_j. We assume that, for fixed j, $\mathcal{W}_k^{(j)}$ are identically distributed for all k, with the same distribution as $\mathcal{W}^{(j)}$. Now define

$$S_l^{(j)} = \sum_{k=1}^{l} \mathcal{W}_k^{(j)},$$

and define the stopping time $K = \min\{l : S_l^{(j)} \geqslant 0 \text{ for all } j = 1, \ldots, N\}$. Then $E(K) < \infty$, and

$$E(S_K^{(j)}) = E(K)\, E(W^{(j)}).$$

For a proof of theorem 1, see Appendix A.

The weights that are simulated by this algorithm will have expectation proportional to the true weights, as required. The unknown normalizing constant causes no complications since it cancels out in the renormalization of the weights. The method using partial sum weights according to this procedure is termed *Wald random-weight IS*.

It is natural to consider the computational cost of applying Wald random-weight IS. Since its application requires K draws, $\{\mathcal{W}_k^{(j)}, 1 \leqslant k \leqslant K\}$, the cost is thus proportional to $E(K)$. However $E(K)$ is increasing in N. Therefore, to provide a robust method which can be used for large particle populations, it is important to understand how rapidly $E(K)$ increases as a function of N.

To give a formal result concerning this, we shall strengthen the moment conditions on $\mathcal{W}_k^{(j)}$, requiring the existence of some exponential moments, which is still a very reasonable assumption. Weaker results are possible under weaker moment constraints. This result is essentially a classical result from large deviation theory (see for example Varadhan (1984)) adapted to our context.

Theorem 2. Suppose that the distributions of $\{\mathcal{W}^{(j)}\}$ admit a stochastic minorant, i.e.

$$P(\mathcal{W}^{(j)} \leqslant x) \leqslant F(x)$$

for all $x \in \mathbb{R}$, $j = 1, 2, \ldots$, with F a proper distribution function. We shall assume that, for some $\varepsilon > 0$, if $X \sim F$ then

$$E\{\exp(sX)\} = \mu(s) < \infty$$

for all $s \in [-\varepsilon, 0]$, and that $E(X) > 0$. Then there is a positive constant a such that, for all N,

$$E(K) \leqslant a \log(N).$$

For a proof of theorem 2, see Appendix B.

In most applications it is reasonable to assume exponential moments for any given $\mathcal{W}^{(j)}$. For example, in the diffusion application that is considered below, $\mathcal{W}^{(j)}$ are IS weights which are based on polynomial functions of Gaussian random variables. In IS applications we would have that $\mathcal{W}^{(j)}$ has the same distribution for all j, in which case the conditions of theorem 2 would

then hold. However, in particle filtering applications, the distributions of $\mathcal{W}^{(j)}$ will vary with j, and there may not be an appropriate stochastic minorant. Extending the above result is possible, but will require specific assumptions on how the distributions of $\mathcal{W}^{(j)}$ vary with j, and is beyond the scope of this paper. However, in Section 4 we look empirically at how $E(K)$ depends on N for particle filtering applications.

3. High frequency particle filtering

Let $Z=(Y, X)$ be a Markov process, where Y is the observable ('the data') and X the unobservable ('the signal') component of the process. We are interested in estimating X at a sequence of times given all available data up to that given time. This is known as the *filtering problem*.

Here we focus on models where both X and Y evolve continuously in time and their dynamics are given by a system of SDEs. Our methodology extends beyond this and also includes cases where Y is partially observed, or other observations schemes. Discussion of these extensions is left until Section 5.

In our set-up Z is a multi-dimensional diffusion process. Exact solution of the filtering problem in this framework is available only when the system is linear in X and Y (Kalman and Bucy, 1961). For non-linear models, the filtering problem is commonly solved by using a collection of Monte Carlo algorithms known as PFs; see Liu and Chen (1998), Doucet *et al.* (2000) or Fearnhead (2008) for a detailed introduction. PFs are based on sequential application of IS (see Section 2). However, application of PFs to the continuous time filtering problem is hindered by the intractability of the transition density of diffusion processes. For certain classes of diffusions unbiased estimators of transition densities are available, and for these models we can use random-weight IS to solve the filtering problem. In the rest of the section we restrict attention to such processes.

Let Z be a multivariate diffusion process satisfying the SDE

$$\mathrm{d}Z_s = -\Sigma \nabla A(Z_s)\,\mathrm{d}s + \sqrt{(2\Sigma)}\mathrm{d}B_s, \qquad s \in [0, T], \quad Z_0 = z, \tag{2}$$

where $A : R^d \to R$ is a potential function, ∇A denotes the vector of d partial derivatives of A, B is a d-dimensional standard Brownian motion and Σ is a symmetric positive definite matrix; $\sqrt{\Sigma}$ is the square root of the matrix given by the Cholesky decomposition. Z is decomposed into $Z=(Y, X)$ with $Y \in R^{d_1}$ and $X \in R^{d_2}$. SDE (2) defines a time reversible process, with invariant density $\exp\{-A(z)\}$, when the latter is integrable in z. We assume that Y is observed at a collection of times $0=t_0<t_1<\ldots<t_n=T$, with observed values y_{t_j}, $j=1, \ldots, n$, $y_{t_i:t_j}$, for $t_i<t_j$, denotes the vector $(y_{t_i}, y_{t_{i+1}}, \ldots, y_{t_j})$. We are particularly interested in cases where n is large compared with T, i.e. high frequency observations.

Compared with a general multivariate diffusion, SDE (2) poses two important restrictions. The first is the constant diffusion matrix Σ. Therefore, the components of the process are coupled solely in the drift. The second is that the drift is driven by the gradient of a potential function. However, the dynamics (2) are standard in the analysis of many physical systems. For example, in the context of molecular dynamics A is a potential energy, describing interactions between components of the system, and the noise term models thermal activation.

Apart from the intractability of diffusion dynamics, we wish to address the problem of obtaining computationally efficient methods for filtering given high frequency observations. The main idea behind our approach is as follows. We consider a sequence of filtering times $0=s_0<s_1<\ldots<s_m=T$, which is a subset of the observation times. The choice of the s_is is made by the user. The choice does effect the performance of the filter (see Section 4), but for high frequency applications they can be chosen such that $m \ll n$ where n is the number of observations.

We design PFs which at each time s_i estimate the filtering distribution of X_{s_i}. Our methods involve simulating a skeleton of (Y, X) at Poisson-distributed times on $[s_{i-1}, s_i]$ according to a simulation smoother and associating each such skeleton with appropriate random weights. A key feature of the method is that the number of points at which we simulate (Y, X) does not increase with the frequency of the observed data. This is achieved by exploiting the particular coupling of the diffusion components in SDE (2) and the characteristics of the unbiased estimators of the transition density.

3.1. Mathematical development

To complete our model we need a prior density for X_0, which we denote by π_0. When SDE (2) is ergodic, a natural choice is the conditional invariant density, $\pi_0(x_0) \propto \exp\{-A(y_0, x_0)\}$. Given π_0 and a set of observations, our aim is to obtain particle approximations to the filtering densities of X at the collection of times $0 = s_0 < s_1 < \ldots < s_m = T$. The ith filtering density is the density of X_{s_i} given $y_{0:s_i}$, and it is denoted by $\pi_{s_i}(x_i)$. Moreover, we denote the joint density of $(X_{s_{i-1}}, X_{s_i})$ given $y_{0:s_i}$ by $\pi_{s_i}(x_{i-1}, x_i)$, and the corresponding conditional density by $\pi_{s_i}(x_i|x_{i-1})$. Clearly, $\pi_{s_i}(x_i)$ is a marginal of $\pi_{s_i}(x_{i-1}, x_i)$. Generally, a density that is superscripted by (N) will denote its particle approximation; for example $\pi_{s_i}^{(N)}$ is a set of N weighted particles, $\{x_i^{(j)}, w_i^{(j)}\}_{j=1}^N$, with $w_{i-1}^{(j)} \geqslant 0$, for all j, approximating $\pi_{s_i}(x_i)$. There are four main steps in the process of deriving the particle approximations. The second and third steps are common to most PFs.

The first step is to express the filtering densities as marginals of distributions defined on the space of paths of the diffusion Z. Let $\mathbb{Q}(y_{s_{i-1}:s_i}, \pi_{s_{i-1}})$ denote the law of Z, conditionally on $Y_{t_j} = y_{t_j}$ for all $s_{i-1} \leqslant t_j \leqslant s_i$, and with initial measure at time s_{i-1}, $X_{s_{i-1}} \sim \pi_{s_{i-1}}$. This is a probability measure on the space of paths of Z on $[s_{i-1}, s_i]$ which are consistent with the data. It follows easily from the Markov property of Z that $\pi_{s_i}(x_{i-1}, x_i)$ is a marginal density obtained from $\mathbb{Q}(y_{s_{i-1}:s_i}, \pi_{s_{i-1}})$ by integrating out the path connecting the end points.

The second step of the approach is to replace $\pi_{s_{i-1}}$ by an existing particle approximation $\pi_{s_{i-1}}^{(N)}$. Then, a particle approximation at time s_i is obtained by sampling from $\mathbb{Q}(y_{s_{i-1}:s_i}, \pi_{s_{i-1}}^{(N)})$ and storing only the x_{s_i}-value. As we cannot sample directly from $\mathbb{Q}(y_{s_{i-1}:s_i}, \pi_{s_{i-1}}^{(N)})$ we use IS. It is easy to obtain a particle approximation to $\pi_0^{(N)}$ by IS. Thus, the scheme that we describe below can be used iteratively to build the approximations for all filtering times, since it shows how to propagate the particle approximations from each time point to the next.

The third step is to devise an appropriate proposal distribution, and corresponding IS weight, to approximate $\mathbb{Q}(y_{s_{i-1}:s_i}, \pi_{s_{i-1}}^{(N)})$ by using IS. We specify the proposal distribution by first choosing the marginal distribution of the end points, $x_{s_{i-1}}, x_{s_i}$, and then selecting a stochastic process for interpolation between the end points. For the former, we follow the generic proposal distribution that was used in the PF of Pitt and Shephard (1999). The proposal is denoted by $\nu_{s_{i-1},s_i}^{(N)}(x_{i-1}, x_i)$ and takes the generic form

$$\nu_{s_{i-1},s_i}^{(N)}(x_{i-1}^{(j)}, x_i) \propto \beta_{i-1}^{(j)}\, q_{s_i}(x_i|x_{i-1}^{(j)}, y_{0:s_i}).$$

Note that the proposal distribution has the same discrete support as $\pi_{s_{i-1}}^{(N)}$ for $x_{s_{i-1}}$, though with potentially different probabilities. Simulation from $\nu_{s_{i-1},s_i}^{(N)}$ is achieved by choosing an existing particle $x_{i-1}^{(j)}$ with probability $\beta_{i-1}^{(j)}$, and then simulating x_i from the density $q_{s_i}(x_i|x_{i-1}^{(j)}, y_{0:s_i})$. The optimal choice of $\nu_{s_{i-1},s_i}^{(N)}$ is discussed in Section 3.4.

For interpolation we use a conditioned Gaussian Markov process. We define the stochastic process $W = (Y, X)$ on $[s_{i-1}, s_i]$, with $Y \in R^{d_1}$, $X \in R^{d_2}$ and

$$W_s = W_{s_{i-1}} + \sqrt{(2\Sigma)} B_{s-s_{i-1}}$$

for $s \geqslant s_{i-1}$, where B is a standard d-dimensional Brownian motion. Let $\mathbb{W}(y_{s_{i-1}:s_i}, \nu^{(N)}_{s_{i-1},s_i})$ denote the law of W conditionally on $Y_{t_j} = y_{t_j}$ for all $s_{i-1} \leqslant t_j \leqslant s_i$, where $(X_{s_{i-1}}, X_{s_i}) \sim \nu_{s_{i-1},s_i}$. $\mathbb{W}(y_{s_{i-1}:s_i}, \nu^{(N)}_{s_{i-1},s_i})$ is a mixed Gaussian law, where the mixing is induced by $\nu^{(N)}_{s_{i-1},s_i}$. The conditional Gaussian law corresponds to what is sometimes called a *simulation smoother* in the literature (see for example de Jong and Shephard (1995)). $\mathbb{W}(y_{s_{i-1}:s_i}, \nu^{(N)}_{s_{i-1},s_i})$ has tractable finite dimensional distributions due to the Markovian structure of W: see Appendix C. We shall use this *mixture of simulation smoothers* as a proposal distribution within IS to sample from $\mathbb{Q}(y_{s_{i-1}:s_i}, \pi^{(N)}_{s_{i-1}})$. The following theorem then gives the likelihood ratio between proposal and target distribution, which by definition is the IS weight.

Theorem 3. Under mild technical and standard conditions (see Appendix D) $\mathbb{Q}(y_{s_{i-1}:s_i}, \pi^{(N)}_{s_{i-1}})$ is absolutely continuous with respect to $\mathbb{W}(y_{s_{i-1}:s_i}, \nu^{(N)}_{s_{i-1},s_i})$, with density given by

$$\frac{w^{(j)}_{i-1}}{\beta^{(j)}_{i-1}} \frac{\mathcal{G}_{s_i - s_{i-1}}(x_i | x^{(j)}_{i-1})}{q_{s_i}(x_i | x^{(j)}_{i-1}, y_{0:s_i})} \exp\left[-\frac{1}{2} \{ A(y_{s_i}, x_i) - A(y_{s_{i-1}}, x^{(j)}_{i-1}) \} \right] \exp\left\{ -\int_{s_{i-1}}^{s_i} \phi(Y_s, X_s)\, \mathrm{d}s \right\}, \tag{3}$$

where ϕ is given in Appendix D, and $\mathcal{G}$ is the conditional density of X_{s_i} given $X_{s_{i-1}}$ and Y_{t_j} for $s_{i-1} \leqslant t_j \leqslant s_i$, when $W = (X, Y)$ is the Gaussian interpolating process that was defined above (see part (b) of proposition 1 in Appendix C). The constant of proportionality is a function of $y_{0:s_i}$.

For a proof of theorem 3, see Appendix D.

Theorem 3 implies that each pair $(x^{(j)}_{i-1}, x_i)$ proposed from $\nu^{(N)}_{s_{i-1},s_i}$ will have to be weighted according to expression (3) to provide a sample from $\pi^{(N)}_{s_i}(x_{i-1}, x_i)$. However, the weight cannot be explicitly computed owing to the last term in the product, which is of the form

$$\exp\left\{ -\int_{s_{i-1}}^{s_i} \phi(Y_s, X_s)\, \mathrm{d}s \right\}. \tag{4}$$

The fourth component of the approach is to replace this term by an unbiased estimator and to resort to random-weight IS. Random weights can be obtained by the following generic method.

3.2. The Poisson estimator for exponential functionals

Let Φ be an unknown quantity, possibly the realization of a random variable. Consider estimating $w = \exp(-\Phi)$. Let $\tilde{\Phi}_j$, $j = 1, 2, \ldots,$ be a collection of independent identically distributed random variables with $\mathbb{E}(\tilde{\Phi}_j | \Phi) = \Phi$. Note that

$$\mathbb{E}\left\{ \prod_{j=1}^{i} (c - \tilde{\Phi}_j) | \Phi \right\} = (c - \Phi)^i,$$

where we define the product $\Pi_{j=1}^{0}$ to be equal to 1. Then, for any $c \in \mathbb{R}, \delta > 0$ we have

$$\begin{aligned} w = \exp(-\Phi) &= \exp(-c) \sum_{i=0}^{\infty} \left(\frac{c - \Phi}{\delta} \right)^i \frac{\delta^i}{i!} \\ &= \exp(-c) \sum_{i=0}^{\infty} \mathbb{E}\left(\prod_{j=1}^{i} \frac{c - \tilde{\Phi}_j}{\delta} \middle| \Phi, c, \delta \right) \frac{\delta^i}{i!} = \exp(\delta - c)\, \mathbb{E}\left(\prod_{j=1}^{\kappa} \frac{c - \tilde{\Phi}_j}{\delta} \middle| \Phi, c, \delta \right) \end{aligned} \tag{5}$$

where $\kappa \sim \text{Poisson}(\delta)$. Thus, we have an unbiased estimator of w given by

$$\mathcal{W} = \exp(\delta - c) \prod_{j=1}^{\kappa} \frac{c - \tilde{\Phi}_j}{\delta}. \tag{6}$$

This is the basic *Poisson estimator* as introduced in Wagner (1987); see Beskos *et al.* (2006) for its use in estimation of diffusions, Fearnhead *et al.* (2008) for extensions and its use in sequential IS and Papaspiliopoulos (2009) for an overview of techniques for unbiased estimation using power series expansions. When w is an IS weight, we shall call $\mathcal{W}$ a *random weight* (see Section 2).

The introduction of auxiliary randomness is not without cost, and the choice of parameters c and δ is critical to the stability of $\mathcal{W}$. This is discussed in detail in Fearnhead *et al.* (2008), who showed that for stability as c increases δ should also increase linearly with c.

In the context of expression (4), $\Phi = \int_{s_{i-1}}^{s_i} \phi(Y_s, X_s)\mathrm{d}s$, and we can take $\tilde{\Phi} = (s_i - s_{i-1})\,\phi(Y_\psi, X_\psi)$, for ψ uniformly distributed on (s_{i-1}, s_i). Thus, we obtain the following unbiased estimator of expression (4):

$$\exp\{(\lambda - r)(s_i - s_{i-1})\}\lambda^{-\kappa} \prod_{j=1}^{\kappa} \{r - \phi(Y_{\psi_j}, X_{\psi_j})\}, \tag{7}$$

where $\kappa \sim \text{Poisson}\{\lambda(s_i - s_{i-1})\}$ and the ψ_js are an ordered uniform sample on (s_{i-1}, s_i). Note that, compared with the general Poisson estimator (6), we have taken $c = r(s_i - s_{i-1})$, and $\delta = \lambda(s_i - s_{i-1})$, for some $\lambda > 0$. Calculating expression (7) involves simulating (Y, X) at κ time points. The distribution of (Y, X) is that of the proposal process $\mathbb{W}$, conditioned on $X_{s_{i-1}} = x_{i-1}^{(j)}$ and $X_{s_i} = x_i$. The pairs (Y_{ψ_j}, X_{ψ_j}) can be simulated sequentially; see part (d) of proposition 1 in Appendix C.

If ϕ is bounded we can choose $r = \max(\phi)$ to ensure that estimator (7), and hence the random weight, is positive. When ϕ is unbounded it is impossible to choose a constant r which will ensure positive random weights. Of course, we can make the probability of negative weights arbitrarily small by choosing r arbitrarily large. However, this comes at a computational cost. As mentioned above, to control the variance of the weights we need to choose $\lambda = O(r)$, and thus the computational cost of estimator (7) is proportional to r. Instead we propose to choose r to be an estimate of the likely maximum of $\phi(Y, X)$ for our proposed path, and then use the Wald identity of Section 2 to ensure positivity of the weights.

3.3. A Wald random-weight particle filter

Having developed the four main components of the filtering algorithm, we can combine them with Wald random weight IS to provide a PF for the continuous time filtering problem under study. In what follows, the random weight of a proposed path (in steps 3(a) and 3(b)) refers to an unbiased estimator of expression (3), composed by exact computation of the first two terms for given end points, and an unbiased estimator of the last term using the Poisson estimator and the simulation smoother. We have the following algorithm.

3.3.1. Wald random weight particle filter

First, simulate $x_0^{(j)}$, $j = 1, \ldots, N$, from ν_{s_0} and weight each value by $w_0^{(j)} = (\mathrm{d}\pi_{s_0}/\mathrm{d}\nu_{s_0})(x_0^{(j)})$. Then for $i = 1, \ldots, m$ carry out the following steps.

Step 1: calculate the effective sample size of the $\{\beta_{i-1}^{(k)}\}_{k=1}^N$, $\text{ESS} := \{\Sigma_{k=1}^N (\beta_{i-1}^{(k)})^2\}^{-1}$.
Step 2: if $\text{ESS} < C$, for some fixed constant C, then for $j = 1, \ldots, n$ simulate $k_{i-1,j}$ from $p(k) \propto \beta_{i-1}^{(k)}$, $k = 1, \ldots, N$, and set $\delta_i^{(j)} = 1$. Otherwise for $j = 1, \ldots, n$ set $k_{i-1,j} = j$ and $\delta_i^{(j)} = \beta_{i-1}^{(j)}$.
Step 3:

(a) for $j = 1, \ldots, n$ simulate $x_i^{(j)}$ from $q_{s_i}(x_i | x_{i-1}^{(k_{i,j})}, y_{0:s_i})$ and generate a random weight $\mathcal{W}_i^{(j)}$;

(b) if $\mathcal{W}_i^{(j)} < 0$ for at least one j, then for $j = 1, \ldots, N$ simulate new random weight $\mathcal{W}_i^{*(j)}$, and set $\mathcal{W}_i^{(j)} = \mathcal{W}_i^{(j)} + \mathcal{W}_i^{*(j)}$. Repeat.

Step 4: assign particle $x_i^{(j)}$ a weight $w_i = \delta_i^{(j)} \mathcal{W}_i^{(j)}$.

Note that our specification of the PF allows for the choice of whether to resample at each time step. This choice is based on the effective sample size of the particle weights (see for example Liu and Chen (1998)). Additionally, step 3(b) generates random weights for all particles until they are all positive. This is according to the Wald random-weight IS of Section 2.

3.4. Implementation

To implement the WRWPF we first need to choose the joint proposal density $\nu_{s_{i-1},s_i}^{(N)}(x_{i-1}^{(j)}, x_i)$ for the current particle at time s_{i-1} and the new particle at time s_i. This density should be easy to simulate from and to provide an approximation to $\pi_{s_i}^{(N)}(x_{i-1}, x_i)$. A simple approach is to approximate the SDE via the Euler discretization. The Euler discretization defines an approximate transition density over time interval $s_i - s_{i-1}$, which can be factorized as $p(y_{s_i}|x_{i-1}, y_{s_{i-1}}) \times p(x_i|y_{s_i}, x_{i-1}, y_{s_{i-1}})$. We then define $\beta_{i-1}^{(j)} = p(y_{s_i}|x_{i-1}^{(j)}, y_{s_{i-1}})$ and $q_{s_i}(x_i|x_{i-1}^{(j)}, y_{0:s_i}) = p(x_i|y_{s_i}, x_{i-1}^{(j)}, y_{s_{i-1}})$. For alternative approaches for designing this proposal distribution for general state space models see Pitt and Shephard (1999).

Secondly we need to choose r and λ for the Poisson estimator. The approach that we took for the study in Section 4 was to choose r dependent on $x_{i-1}^{(j)}$ and x_i, the start and end points of the path. Given these end points, and the data, we can estimate an interval for each component of (Y, X), that the proposal path will lie in with high probability. For high frequency data the interval for each component of Y will be close to the minimum and maximum of the observations in $y_{s_{i-1}:s_i}$. For the X-components we use the fact that the proposal process is a Gaussian process with known mean and covariance to obtain appropriate intervals. We set r to an estimate of the maximum value of $\phi(Y, X)$ under the assumption that each component of the path lies within its respective interval. Once we have chosen r we use the guidelines in Fearnhead *et al.* (2008) to choose λ.

The advantage of using the Wald identity to produce positive random weights is that on the basis of the above recipe we can construct relatively crude estimates for r, which are quick to calculate. If we do obtain a poor value for r, then it affects only the computational cost within step 3(b) of the WRWPF. By comparison, methods based on, for example, truncating negative weights can introduce a large bias if we choose too small a value for r.

4. Illustration of the methodology

We illustrate our methods on the following two examples. The first is a linear SDE for which the exact filtering densities can be calculated by using the Kalman filter. We use this example to evaluate the performance of our method and compare it with filters that require discretization. We then consider a model problem that is taken from molecular dynamics based on a double-well potential (see for example Metzner *et al.* (2006)).

In both cases we implemented the WRWPF, using the stopping time idea of theorem 1 to correct for negative weights. In all cases we used 1000 particles. Resampling for the PFs was via the stratified sampling approach of Carpenter *et al.* (1999) and was used when the effective sample size of the particle weights dropped below $N/2$. For simplicity we have considered two-dimensional systems with $\Sigma = \mathrm{diag}\{1/(2\varepsilon), \frac{1}{2}\}$. In this setting the value of ε governs the relative speed of the observed and unobserved processes, and we investigate the performance of our

method for various values of ε. See Givon *et al.* (2009) for alternative PF methods for such problems.

4.1. Example 1: Ornstein–Uhlenbeck process

Taking $A(u) = (u - \mu)^* \, Q(u - \mu)/2, u, \mu \in R^d$, gives rise to a subset of the family of Ornstein–Uhlenbeck processes. If Q is a symmetric positive definite matrix, Z is ergodic with Gaussian invariant law with mean μ and inverse covariance matrix Q. We take $d = 2$, set $Q_{11} = Q_{22} = 1$ and $Q_{12} = Q_{21} = -0.9$, and without loss of generality $\mu = (0, 0)$. This produces a process with a symmetric stationary distribution, with correlation of 0.9 between the two components. An example realization of the process is given in Fig. 1 with $\varepsilon = 1/100$. We then applied the WRWPF to analyse these data, using 10^5 observations. We chose 100 equally spaced filtering times. The filtered mean and regions of ± 2 standard deviations are shown in Fig. 1 together with the exact quantities as calculated by the Kalman filter. By eye, there is no noticeable difference between the Kalman filter and WRWPF results.

We also compared the performance of the WRWPF with a filter based on discretizing time. For this latter filter, given a set of filtering times inference is performed based just on the observations at those times. The state dynamics between filtering times are approximated through an Euler approximation to the SDE. A standard PF can then be applied to the resulting discrete time problem; in practice we implemented a fully adapted version of the PF of Pitt and Shephard (1999). We implemented such a particle filter with 1000 particles, which we call a discrete time particle filter (DTPF). For this problem, after discretizing time we have a simple linear Gaussian state space model, for which we can calculate the exact filtering distributions by using the

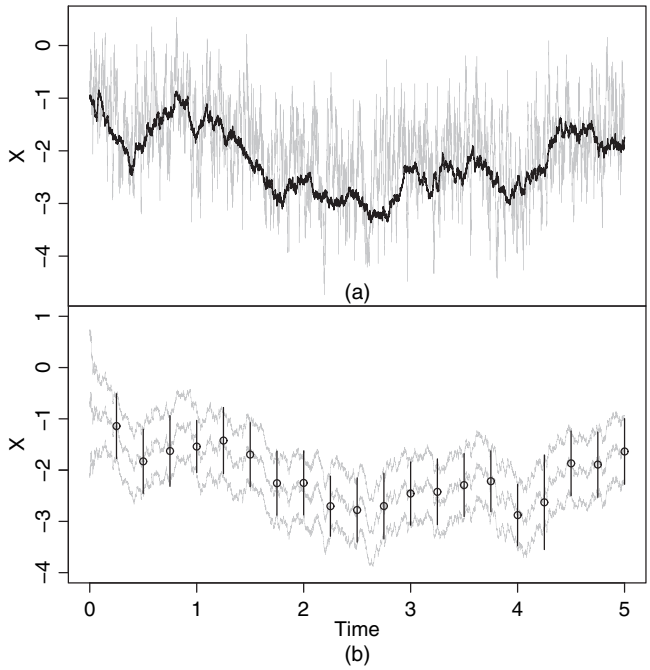

Fig. 1. (a) Simulated realization of the Ornstein–Uhlenbeck process (———, (slow) unobserved state; ———, (fast) observed state) and (b) posterior mean and intervals of two posterior standard deviations away from the mean, computed exactly by using the Kalman filter (———) (shown at every observation time) and estimated by the WRWPF (ϕ) (shown at 20 observation times) based on 100 filtering times

Kalman filter. We also looked at this approach, which we denote the discrete time Kalman filter (DTKF). The DTKF is equivalent to the performance of the DTPF with an infinite number of particles.

A comparison of the three methods is shown in Fig. 2, for various numbers of filtering times and various values of ε. We plot the mean-square error MSE between each filter's estimate of the mean of the filtering distribution, and the exact filtering mean. Note that the effect of ε on the results is small, except in terms of the best choice of the frequency of filtering times, with this reducing as ε increases.

The WRWPF gives a substantial reduction in mean-square error over the other two filters. Furthermore we see that the Monte Carlo error within the particle filter is small, as both the DTPF and the DTKF give almost identical results. The WRWPF's performance is also robust to the number of filtering times, as it uses all the information in the data regardless, unlike the DTPF or DTKF. Its performance is slightly worse for smaller numbers of filtering times, owing to the increased Monte Carlo variation in simulating the weights in these cases. The computational cost of the WRWPF is reasonably robust to the choice of the number of filtering times.

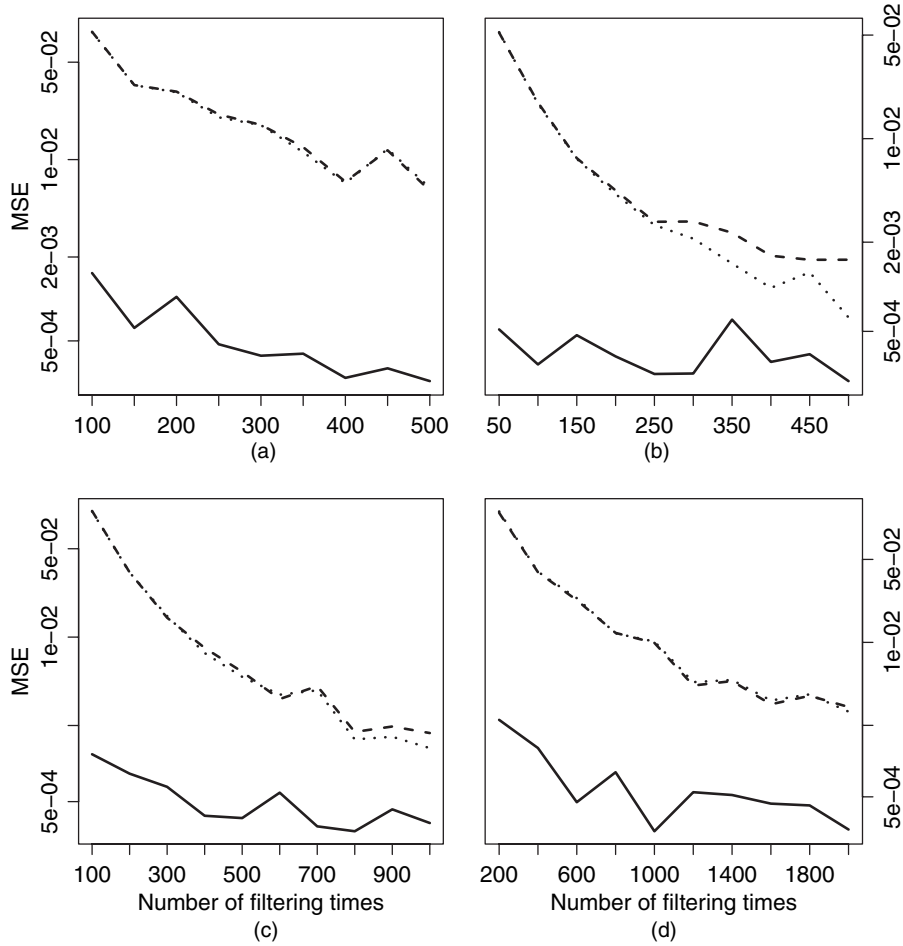

Fig. 2. Comparison of the WRWPF (———), DTPF (- - - - - - -) and DTKF (· · · · · · ·) at approximating the filtered mean: the results are for the mean-square error relative to the truth, and for (a) $\varepsilon = 1/10$, (b) $\varepsilon = 1/25$, (c) $\varepsilon = 1/100$ and (d) $\varepsilon = 1/400$

For example, for $\varepsilon = 1/100$ the total number of simulations per particle (which is equal to the number of filtering steps plus the number of points simulated in calculating the weights) ranges from 800 (300 filtering times) to 1250 (1000 filtering times) over the different choices, though it would start to increase linearly as the number of filtering times increases beyond 1000.

The cost of analysing the whole data set with our new algorithm with 1000 particles is of the order of a few minutes (programmed in R). The WRWPF has the advantage over the DTPF of unbiased weights, but at the cost of an increased variability (the WRWPF performs resampling roughly 10 times as often as the DTPF). However, the results from our study compare both methods on the basis of their mean-square error—and the extra variability of the WRWPF's weights is small compared with the increased accuracy through using unbiased weights.

Note that a direct comparison of WRWPF and DTPF for the same number of filtering times is unfair—as the amount of simulation per particle for the DTPF is just equal to the number of filtering times. However, even taking this into account (for $\varepsilon = 1/100$ compare the WRWPF with 300 times *versus* the DTPF with 800 filtering times) the WRWPF is substantially more accurate. The WRWPF is also more accurate if we allow the DTPF to have more particles. The WRWPF outperforms the DTKF, and the DTKF gives the limiting performance of the DTPF as the number of particles tends to ∞.

We also investigate the effect that the number of particles has on the computational efficiency of the WRWPF. As mentioned above, increasing the number of particles will increase the amount of simulation per particle required to ensure positive weights. To investigate this empirically we reanalysed the Ornstein–Uhlenbeck model with $\varepsilon = 1/100$ under the WRWPF with various numbers of particles. The results suggest that the amount of simulation per particle

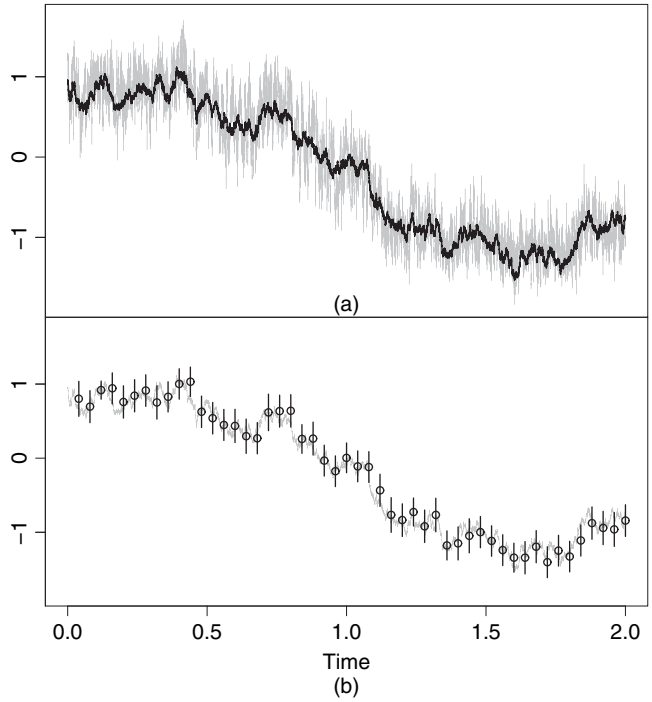

Fig. 3. Results for the double-well model (the WRWPF was run for 500 filtering times, but for clarity results at only 50 are shown): (a) true state (———) and observed process (———); (b) true state (———), filtered means (○) and regions ±2 standard errors (|)

increases with the logarithm of the number of particles, increasing by roughly 150 each time the number of particles doubles. This is consistent with theorem 2.

4.2. Example 2: double-well potential

A more challenging example is specified by the following potential function, where we take $d_1 = d_2 = 1$, and

$$A(y, x) = q_1(y^2 - \mu)^2 + q_2(y - q_3 x)^2, \qquad q_1, q_2 > 0, \mu, q_3 \in \mathbf{R}. \tag{8}$$

The potential produces a stationary distribution with two modes, at $(y, x) = (1, 1)$ and $(y, x) = (-1, -1)$. We set $\varepsilon = 1/100$.

Here we focus solely on the performance of the WRWPF. We simulated data over two units of time, with 2×10^5 observations. Our simulated data were chosen to include a transition of the process between the two modes. We analysed the data by using 500 filtering times. The results are shown in Fig. 3. In this example we simulated the process at an average of eight Poisson time points between each time point, which suggests that having more frequent filtering times would be preferable. However, even in this case, resampling was required only at every other time step.

5. Discussion

In this paper we have introduced the idea of using the Wald identity to ensure positivity of random IS weights and developed filtering methods for high frequency observations whose computational cost is independent of the observation frequency. Although we have implemented the two ideas together, there is scope for implementing each independently.

Our high frequency filtering application focused on models for which we can do the filtering without introducing time discretization approximations. The practical advantage of such an approach was discussed in Fearnhead *et al.* (2008). Although we considered only partial observations from the SDE model (2), our method can be extended to models with other observations schemes (see those discussed in Fearnhead *et al.* (2008)), and also to models where the drift of the SDE is of the form $-\Sigma\nabla A(Z_s) + DZ_s$, where D is a $d \times d$ matrix. (For the latter extension we need to redefine the proposal process W_t so that it is the solution of an SDE with drift DW_s.)

The idea of calculating the filtering distributions at a subset of the observation times, but in a way that still uses all the information in the data, can be applied even more generally. For example we could analyse models where the instantaneous variance is not constant. In this case our proposal process W_t must have the same non-constant instantaneous variance, and thus simulation of the path in between filtering times must be done from an approximation to the true diffusion process. Although introducing time discretization error, such an approach will still keep the computational savings, with computational cost being independent of the frequency of the observations. There is similar scope to apply the filtering idea to discrete time models.

Wald random-weight IS can also be applied more generally than considered here. For example it is simple to apply this idea to ensure positivity of weights for the models that were considered in Fearnhead *et al.* (2008). The alternative approach that was used in Fearnhead *et al.* (2008) to ensure positive weights was based around a method for simulating the skeleton of a path of Brownian motion together with bounds on the path (Beskos *et al.*, 2008). The advantage of using the Wald identity is primarily simplicity of implementation. An important area of future research is to investigate whether existing convergence results (as $N \to \infty$) for PFs apply directly to WRWPFs.

Acknowledgements

We thank the Joint Editor, Associate Editor and the referees for many helpful suggestions, and Pierre Del Moral, Dan Crisan and Terry Lyons for motivating discussions. The second author acknowledges financial support by the Spanish Government through a 'Ramon y Cajal' fellowship and grant MTM2008-06660.

Appendix A: Proof of theorem 1

Theorem 1 is essentially Wald's identity (see for example proposition 2.18 of Siegmund (1985)). All we need to do is to verify the finiteness of $E(K)$. However, by a standard expansion

$$E\{(S_l^{(j)} - lw_j)^4\} = lv_j + 6l(l-1)\sigma_j^4$$

where $v_j = E\{(\mathcal{W}_k^{(j)} - w_j)^4\}$ and $\sigma_j^2 = E\{(\mathcal{W}_k^{(j)} - w_j)^2\}$. Now set $\sigma^2 = \max_j\{\sigma_j^2\}$, $v = \max_j\{v_j\}$ and $w = \min_j\{w_j\}$ so that

$$\begin{aligned} P(K > r) &\leqslant P\left\{\bigcup_{j=1}^{N}(S_r^{(j)} < 0)\right\} \leqslant \sum_{j=1}^{N} P(S_r^{(j)} < 0) \\ &\leqslant \sum_{j=1}^{N} P(|S_r^{(j)} - rw_j| > rw_j) \leqslant \sum_{j=1}^{N} \frac{E\{(S_r^{(j)} - rw_j)^4\}}{r^4 w_j^4} \\ &\leqslant \sum_{j=1}^{N} \frac{rv_j + 6r(r-1)\sigma_j^4}{r^4 w_j^4} \leqslant \frac{Nv}{r^3 w^4} + \frac{6N\sigma^4}{r^2 w^4} = O(r^{-2}). \end{aligned} \tag{9}$$

Since this expression has a finite sum (in r) this implies the finiteness of $E(K)$ as required.

Appendix B: Proof of theorem 2

By Markov's inequality, for $-\varepsilon \leqslant s \leqslant 0$,

$$P(S_k^{(j)} < 0) \leqslant \mu(s)^k.$$

Now, from the conditions imposed on μ, it is continuously differentiable with $\mu'(0) = E(X) > 0$. Therefore for some $s < 0$ sufficiently close to 0, s^* say, $\mu(s^*) < 1$.

$$P(\cup_{j=1}^{N} S_k^{(j)} < 0) \leqslant N\,\mu(s^*)^k.$$

Therefore

$$P(K > r) \leqslant \min\{1, N\,\mu(s^*)^r\}.$$

Now set

$$c = \left\lfloor \frac{-\log(N)}{\log\{\mu(s^*)\}} \right\rfloor + 1$$

(where $\lfloor\cdot\rfloor$ denotes the integer part) so that

$$P(K > r + c) \leqslant \min\{1, \mu(s^*)^r\}.$$

Then

$$\begin{aligned} E(K) &= \sum_{r=0}^{\infty} P(K > r) \\ &\leqslant \sum_{r=0}^{c} P(K > r) + \sum_{r=c+1}^{\infty} P(K > r) \\ &\leqslant c + 1 + \sum_{i=1}^{\infty} \mu(s^*)^i \\ &= O\{\log(N)\} \end{aligned}$$

as required.

Appendix C: Properties of *W*-process

Proposition 1. Consider the process $W=(Y,X)$ that was defined in Section 3, and an arbitrary collection of times $t_0<t_1<\ldots<t_n$. Assume that Σ is blocked in terms of Σ_1 (corresponding to instantaneous variance of Y), Σ_2 (corresponding to X) and Σ_{12} (corresponding to the instantaneous covariance of Y with X). Then, we have the following properties.

(a) The joint distribution of $(Y_{t_1},\ldots,Y_{t_n})$ conditionally on $Y_{t_0}=y_{t_0}$ is independent of X_{t_0}, and it is given by the Markov transitions $Y_{t_i}|Y_{t_{i-1}}\sim N\{Y_{t_{i-1}},(t_i-t_{i-1})\Sigma_1\}$.

(b) For any $l=1,\ldots,n$, the distribution of X_{t_l} conditionally on $X_{t_0}=x_0$ and $Y_{t_i}=y_{t_i}, 0\leqslant i\leqslant l$, is independent of Y_s for any $s>t_l$, and it has a Gaussian distribution,

$$N\{x_0+\Sigma_{12}^*\Sigma_1^{-1}(y_{t_l}-y_{t_0}),2(\Sigma_2-\Sigma_{12}^*\Sigma_1^{-1}\Sigma_{12})(t_l-t_0)\},$$

evaluated at a point $x_1\in R^{d_2}$, by $\mathcal{G}_{t_l-t_0}(x_1|x_0)$.

(c) For any $l=1,\ldots,n-1$, the conditional distribution of X_{t_l} given $X_{t_0}=x_0$, $X_{t_n}=x_n$ and $Y_{t_i}=y_{t_i}, 0\leqslant i\leqslant n$, is Gaussian, with mean and variance respectively

$$\begin{gathered}\frac{t_n-t_l}{t_n-t_0}\{x_0+\Sigma_{12}^*\Sigma_1^{-1}(y_{t_l}-y_{t_0})\}+\frac{t_l-t_0}{t_n-t_0}\{x_n+\Sigma_{12}^*\Sigma_1^{-1}(y_{t_l}-y_{t_n})\},\\ 2\frac{(t_n-t_l)(t_l-t_0)}{t_n-t_0}(\Sigma_2-\Sigma_{12}^*\Sigma_1^{-1}\Sigma_{12}).\end{gathered}\tag{10}$$

(d) Assume that $Y_{t_i}=y_{t_i}, 0\leqslant i\leqslant n$, and that X has also been observed at two time points, $X_0=x_0$ and $X_{l_k}=x_k$, for some $k\leqslant n$. Consider a time $t_{l-1}<s<t_l$, for some $1<l\leqslant k$. Then, conditionally on all observed values, the law of (Y_s,X_s) can be decomposed as follows. Y_s is distributed according to a Gaussian distribution,

$$N\left\{\frac{s-t_{l-1}}{t_l-t_{l-1}}y_{t_l}+\frac{t_l-s}{t_l-t_{l-1}}y_{t_{l-1}},\frac{(s-t_{l-1})(t_l-s)}{t_l-t_{l-1}}\Sigma_1\right\},$$

specified in property (c).

The use of properties (b)–(d) is for simulating values of the proposed path at arbitrary times. Property (b) gives the conditional distribution of the X-process at a time point where an observation has been made, and properties (c) and (d) give the conditional distribution of (Y,X) at an arbitrary, non-observation, time.

These equations simplify substantially when $\Sigma_{12}=0$, in which case the proposal distributions for the X- and Y-processes are independent.

Appendix D: Proof of theorem 3

The mild regularity conditions that are referred to in the statement of theorem 3 are those that permit Girsanov's formula to hold. A particularly useful and weak set of conditions to ensure this are given in Rydberg (1997). In the case of unit-diffusion-coefficient diffusions on R^d with locally bounded drift functions, these conditions boil down to just requiring non-explosivity of the SDE satisfying expression (2).

Girsanov's formula gives an expression for the density of the law of the diffusion sample path with respect to the appropriate Brownian dominating measure, on an interval conditional on the left-hand end points. The log-density is given by

$$-\frac{1}{2}\int_0^{\dagger}\nabla A(Z_s)^*\,\mathrm{d}Z_s-\frac{1}{4}\int_0^*\nabla A(Z_s)^*\Sigma\nabla A(Z_s)\,\mathrm{d}s,\tag{11}$$

where the asterisk denotes the Euclidean inner product. Assuming that ∇A is continuously differentiable in all its arguments, we apply Itô's formula to eliminate the stochastic integral and rewrite the log-density as in expression (3), where

$$\phi(u):=\frac{1}{4}\left\{\nabla A(u)^*\Sigma\nabla A(u)-2\sum_{i,j}\Sigma_{ij}\frac{\partial^2}{\partial u_i\,\partial u_j}A(u)\right\}.\tag{12}$$

References

Andrieu, C., Doucet, A. and Holenstein, R. (2010) Particle Markov chain Monte Carlo methods (with discussion). *J. R. Statist. Soc.* B, **72**, 269–342.

Andrieu, C. and Roberts, G. (2009) The pseudo-marginal approach for efficient monte carlo computations. *Ann. Statist.*, **37**, 697–725.

Barndorff-Nielsen, O. E. and Shephard, N. (2001) Non-Gaussian Ornstein–Uhlenbeck-based models and some of their uses in financial economics (with discussion). *J. R. Statist. Soc.* B, **63**, 167–241.

Beskos, A., Papaspiliopoulos, O. and Roberts, G. O. (2008) A factorisation of diffusion measure and finite sample path constructions. *Methodol. Comput. Appl. Probab.*, **10**, 85–104.

Beskos, A., Papaspiliopoulos, O., Roberts, G. O. and Fearnhead, P. (2006) Exact and computationally efficient likelihood-based estimation for discretely observed diffusion processes (with discussion). *J. R. Statist. Soc.* B, **68**, 333–382.

Carpenter, J., Clifford, P. and Fearnhead, P. (1999) An improved particle filter for non-linear problems. *IEE Proc. Radar Sonar Navign*, **146**, 2–7.

Del Moral, P. and Miclo, L. (2000) A Moran particle system approximation of Feynman-Kac formulae. *Stoch. Processes Appl.*, **86**, 193–216.

Doucet, A., Godsill, S. and Andrieu, C. (2000) On sequential Monte Carlo sampling methods for Bayesian filtering. *Statist. Comput.*, **10**, 197–208.

Fearnhead, P. (2008) Computational methods for complex stochastic systems: a review of some alternatives to MCMC. *Statist. Comput.*, **18**, 151–171.

Fearnhead, P., Papaspiliopoulos, O. and Roberts, G. O. (2008) Particle filters for partially observed diffusions. *J. R. Statist. Soc.* B, **70**, 755–777.

Givon, D., Stinis, P. and Weare, J. (2009) Variance reduction for particle filters of systems with time scale separation. *IEEE Trans. Signal Process.*, **57**, 424–435.

Glasserman, P. and Staum, J. (2001) Conditioning on one-step survival for barrier options. *Ops Res.*, **49**, 923–937.

de Jong, P. and Shephard, N. (1995) The simulation smoother for time series models. *Biometrika*, **82**, 339–350.

Kalman, R. and Bucy, R. (1961) New results in linear filtering and prediction theory. *J. Basic Engng*, **83**, 95–108.

Liu, J. S. (2001) *Monte Carlo Strategies in Scientific Computing*. New York: Springer.

Liu, J. S. and Chen, R. (1998) Sequential Monte Carlo methods for dynamic systems. *J. Am. Statist. Ass.*, **93**, 1032–1044.

Metzner, P., Schütte, C. and Vanden-Eijnden, E. (2006) Illustration of transition path theory on a collection of simple examples. *J. Chem. Phys.*, **125**, 084110.

Papaspiliopoulos, O. (2009) A methodological framework for monte carlo probabilistic inference for diffusion processes. In *Inference and Learning in Dynamic Models*. Cambridge: Cambridge University Press.

Pitt, M. K. and Shephard, N. (1999) Filtering via simulation: auxiliary particle filters. *J. Am. Statist. Ass.*, **94**, 590–599.

R Development Core Team (2009) *R: a Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing.

Rousset, M. and Doucet, A. (2006) Discussion on 'Exact and computationally efficient likelihood-based inference for discretely observed diffusion processes' (by A. Beskos, O. Papaspiliopoulos, G. O. Roberts and P. Fearnhead). *J. R. Statist. Soc.* B, **68**, 375–376.

Rydberg, T. H. (1997) A note on the existence of unique equivalent martingale measures in a Markovian setting. *Finan. Stochast.*, **1**, 251–257.

Siegmund, D. (1985) *Sequential Analysis*. New York: Springer.

Varadhan, S. (1984) *Large Deviations and Applications*. New York: Springer.

Wagner, W. (1987) Unbiased Monte Carlo evaluation of certain functional integrals. *J. Comput. Phys.*, **71**, 21–33.