Particle Filters for Partially Observed Diffusions

Paul Fearnhead*, Omiros Papaspiliopoulos† and Gareth O. Roberts†

October 4, 2007

Abstract

In this paper we introduce a novel particle filter scheme for a class of partially-observed multivariate diffusions. We consider a variety of observation schemes, including diffusion observed with error, observation of a subset of the components of the multivariate diffusion and arrival times of a Poisson process whose intensity is a known function of the diffusion (Cox process). Unlike currently available methods, our particle filters do not require approximations of the transition and/or the observation density using time-discretisations. Instead, they build on recent methodology for the exact simulation of the diffusion process and the unbiased estimation of the transition density as described in Beskos et al. (2006b). We introduce the Generalised Poisson Estimator, which generalises the Poisson Estimator of Beskos et al. (2006b). A central limit theorem is given for our particle filter scheme.

Keywords : Continuous-time particle filtering, Exact Algorithm, Auxiliary Variables, Central Limit Theorem, Cox Process

1 Introduction

There is considerable interest in using diffusion processes to model continuous-time phenomena in many diverse scientific disciplines. These processes can be used to model directly the observed data and/or to describe unobserved processes in a hierarchical model. This paper focuses on estimating the path of the diffusion given partial information about it. We develop novel particle filters for analysing a class of multivariate diffusions which are partially observed at a set of discrete time-points.

*Department of Mathematics and Statistics, Lancaster University, U.K., email: p.fearnhead@lancaster.ac.uk

†Department of Statistics, Warwick University, U.K., email: O.Papaspiliopoulos@warwick.ac.uk, Gareth.O.Roberts@warwick.ac.uk

The authors acknowledge helpful comments from the editor and referees.

Particle filtering methods are standard Monte-Carlo methods for analysing partially-observed discrete-time dynamic models (Doucet et al., 2001). They involve estimating the filtering densities of interest by a swarm of weighted particles. The approximation error decreases as the number of particles, N, increases. However, filtering for diffusion processes is significantly harder than for discrete-time Markov models since the transition density of the diffusion is unavailable in all but a few special cases. In many contexts even the observation density is intractable. Therefore, the standard propagation/weighting/re-sampling steps in the particle filter algorithm cannot be routinely applied.

To circumvent these complications, a further approximation, based on a time-discretisation of the diffusion, has been suggested (see for example Crisan et al., 1999; Del Moral et al., 2001). The propagation of each particle from one observation time to the next is done by splitting the time increment into M, say, pieces and performing M intermediate simulations according to an appropriate Gaussian distribution. As M gets large this Gaussian approximation converges to the true diffusion dynamics. In this framework the computational cost of the algorithm is of order $M \times N$, and the true filtering distributions are obtained as both M and N increase.

Our approach does not rely on time-discretisation, but builds on recent work on the Exact Algorithm for the simulation of diffusions (Beskos and Roberts, 2005; Beskos et al., 2006a, 2005b) and on the unbiased estimation of the diffusion transition density (Beskos et al., 2006b, 2005a). This algorithm can be used in a variety of ways to avoid time discretisations in the filtering problem. The potential of the Exact Algorithm in the filtering problem was brought up in the discussion of Beskos et al. (2006b), see the contributions by Chopin, Künsch, and in particular Rousset and Doucet who also suggest the use of a random weight particle filter in this context.

One possibility is simply to use the Exact Algorithm to propagate the particles in the implementation of the Gordon et al. (1993) bootstrap particle filter, thus avoiding entirely the M intermediate approximate simulations between each pair of observation times. We call this the Exact Propagation Particle Filter (EPPF). Where possible, a better approach is to adapt the Exact Algorithm to simulate directly from (a particle approximation to) the filtering density using rejection sampling; we term this the Exact Simulation Particle Filter (ESPF).

However, our favoured method goes in a different direction. We work in the framework of the auxiliary particle filter of Pitt and Shephard (1999), where particles are propagated from each observation time to the next according to a user-specified density and then are appropriately weighted to provide a consistent estimator of the new filtering distribution. Due to the transition density being unavailable, the weights associated with each particle are intractable. However, our approach is to assign to each particle a random positive weight which is an unbiased estimator of the true weight. We call this the Random Weight Particle Filter (RWPF). Our algorithm yields consistent estimates of the filtering distributions. The replacement of the weights in a particle filter by positive unbiased estimators is an interesting possibility in more general contexts than the one considered in this paper.

Indeed, in Section 3.2 we show that this approach amounts to a convenient augmentation of the state with auxiliary variables.

The construction of the unbiased estimators of the weights is one of the main contributions of this paper, and it is of independent interest. This is based on an extension of the *Poisson Estimator* of Beskos et al. (2006b), which we call the *Generalised Poisson Estimator*. This estimator is guaranteed to return positive estimates (unlike the Poisson Estimator) and its efficiency (in terms of variance and computational cost) can be up to orders of magnitude better than the Poisson Estimator. Optimal implementation of the Poisson and the Generalised Poisson estimators is thoroughly investigated theoretically and via simulation.

All three time-discretisation-free particle filters we introduce are easy to implement, with the RWPF being the easiest and the most flexible to adapt to contexts more general than those considered here. A simulation study is carried out which shows that the RWPF is considerably more efficient than the ESPF which is more efficient than the EPPF. We also provide a theoretical result which shows that our filters can have significant computational advantages over time-discretisation methods. We establish a Central Limit Theorem (CLT) for the estimation of expectations of the filtering distributions using either of the EPPF, ESPF and the RWPF. This is an extension of the results of Chopin (2004). The CLT shows that, for a fixed computational cost K, the errors in the particle approximation of the filtering distributions decrease as $K^{-1/2}$ in our methods, whereas it is known that the rate is $K^{-1/3}$ or slower in time-discretisation methods.

The main limitation of the methodology presented here is the requirement that the stochastic differential equation specifying the underlying diffusion process can be transformed to one with orthogonal diffusion matrix, and gradient drift. Although this framework excludes some important model types (such as stochastic volatility models) it incorporates a wide range of processes which can model successfully many physical processes. On the other hand, our methods can handle a variety of discrete-time observation schemes. In this paper we consider three schemes: noisy observations of a diffusion process, observation of a subset of the components of a multivariate diffusion, and arrival times of a Poisson process whose intensity is stochastic and it is given by a known function of a diffusion.

The paper is organised as follows. Section 2 introduces the model for the underlying diffusion and the necessary notation, the observation schemes we consider and the simulated data sets on which we test our proposed methods. Section 3 introduces the RWPF and states the CLT. Section 4 introduces the main tool required in constructing the RWPF, the Generalised Poisson Estimator (GPE). Several theoretical results are established for the GPE, and a simulation study is performed to assess its performance. Section 5 is devoted in the empirical investigation of the performance of the different particle filters we introduce. Several implementation issues are also discussed. Section 6 closes with a discussion on extensions of the methodology and the appendices contain technical results and proofs.

2 Signal, data and assumptions

Let the signal be modelled by a d-dimensional diffusion process

$$\mathrm{d}\mathbf{X}_s = \boldsymbol{\alpha}(\mathbf{X}_s)\,\mathrm{d}s + \mathrm{d}\mathbf{B}_s\ , \quad s \in [0,t]\,. \tag{1}$$

We assume throughout the paper that the drift is known. Our approach requires some assumptions which we summarize in this paragraph: i) $\boldsymbol{\alpha}$ is continuously differentiable in all its arguments, ii) there exists a function $A : \mathbf{R}^d \to \mathbf{R}$ such that $\boldsymbol{\alpha}(\mathbf{u}) = \nabla A(\mathbf{u})$, and iii) there exists $l > -\infty$ such that $\phi(\mathbf{u}) := \big(\|\boldsymbol{\alpha}(\mathbf{u})\|^2 + \nabla^2 A(\mathbf{u})\big)/2 - l \geq 0$. Among these last three conditions i) and iii) are weak and the strictest is ii), which in the ergodic case corresponds to $\mathbf{X}$ being a time-reversible diffusion.

The transition density of (1) is typically intractable but a useful expression is available (see for example Beskos et al., 2006b; Dacunha-Castelle and Florens-Zmirou, 1986)

$$p_t(\mathbf{x}_t \mid \mathbf{x}_0) = \mathcal{N}_t(\mathbf{x}_t - \mathbf{x}_0)\exp\{A(\mathbf{x}_t) - A(\mathbf{x}_0) - lt\}\mathbb{E}\left[\exp\left\{-\int_0^t \phi(\mathbf{W}_s)\mathrm{d}s\right\}\right]\,. \tag{2}$$

In this expression $\mathcal{N}_t(\mathbf{u})$ denotes the density of the d-dimensional normal distribution with mean $\mathbf{0}$ and variance $t\boldsymbol{I}_d$ evaluated at $\mathbf{u} \in \mathbf{R}^d$, and the expectation is taken w.r.t. a Brownian bridge, $\mathbf{W}_s, s \in [0,t]$, with $\mathbf{W}_0 = \mathbf{x}_0$ and $\mathbf{W}_t = \mathbf{x}_t$. Note that the expectation in this formula typically cannot be evaluated.

The data consist of partial observations $y_1, y_2, \ldots, y_n$, at discrete time-points $0 \leq t_1 < t_2 < \cdots < t_n$. We consider three possible observation regimes:

(A) *Diffusion observed with error.* The observation y_i, is related to the signal at time t_i via a known density function $f(y_i|\mathbf{x}_{t_i})$. This model extends the general state-space model by allowing the signal to evolve continuously in time. There is a wide range of applications which fit in this framework, see Doucet et al. (2001) for references.

(B) *Partial Information.* At time t_i we observe $y_i = \zeta(\mathbf{X}_{t_i})$ for some non-invertible known function $\zeta(\cdot)$. For example we may observe a single component of the d-dimensional diffusion. In this model type $f(y_i|\mathbf{x}_{t_i}) = 1$ for all $\mathbf{x}_{t_i}$ for which $\zeta(\mathbf{x}_{t_i}) = y_i$.

(C) *Cox Process.* In this regime the data consist of the observation times t_i which are random and are assumed to be the arrivals of a Poisson process of rate $\nu(\mathbf{X}_s)$, for some known function ν. Such models are popular in insurance (Dassios and Jang, 2005) and finance (Engel, 2000; Duffie and Singleton, 1999), and they have recently been used to analyse data from single molecule experiments (Kou et al., 2005).There is a significant difference between this observation regime and the two previous ones. To have notation consistent with (A) and (B) we let $y_i = t_i$ denote the time of the ith observation; and define the likelihood $f(y_i \mid \mathbf{x}_{t_{i-1}}, \mathbf{x}_{t_i})$ to be the probability density

that the next observation after t_{i-1} is at time t_i. This density can be obtained by integrating

$$\nu(\mathbf{X}_s)\exp\left\{-\int_{t_{i-1}}^{t_i}\nu(\mathbf{X}_s)\mathrm{d}s\right\}, \tag{3}$$

w.r.t. the distribution of $(\mathbf{X}_s, s\in(t_{i-1},t_i))$ conditionally on $\mathbf{X}_{t_{i-1}}=\mathbf{x}_{t_{i-1}}, \mathbf{X}_{t_i}=\mathbf{x}_{t_i}$. The distribution of this conditioned process has a known density w.r.t. the Brownian bridge measure and it is given in Lemma 1 of Beskos et al. (2006b). We can thus show that the density of interest is

$$\frac{\nu(\mathbf{x}_{t_i})\mathcal{N}_{t_i-t_{i-1}}(\mathbf{x}_{t_i}-\mathbf{x}_{t_{i-1}})}{p_{t_i-t_{i-1}}(\mathbf{x}_{t_i}\mid\mathbf{x}_{t_{i-1}})}\exp\{A(\mathbf{x}_{t_i})-A(\mathbf{x}_{t_{i-1}})\}\mathbb{E}\left[\exp\left\{-\int_{t_{i-1}}^{t_i}(\phi(\mathbf{W}_s)+\nu(\mathbf{W}_s))\mathrm{d}s\right\}\right], \tag{4}$$

where expectation is with respect to the law of a Brownian Bridge from $\mathbf{x}_{t_{i-1}}$ to $\mathbf{x}_{t_i}$.

We take a Bayesian approach, and assume a prior distribution for $\mathbf{X}_0$. Our interest lies in the online calculation of the filtering densities, the posterior densities of the signal at time t_i given the observations up to time t_i, for each $1\le i\le n$. While these densities are intractable, we propose a particle filter scheme to estimate recursively these densities at each observation time-point. As we point out in Section 6, our approach allows the estimation of the filtering distribution of the continuous time path $(\mathbf{X}_s, t_{i-1}<s<t_i)$.

A more flexible model for the signal is a diffusion process $\mathbf{Z}$ which solves a more general SDE than the one we have assumed in (1):

$$\mathrm{d}\mathbf{Z}_s=\boldsymbol{b}(\mathbf{Z}_s)\,\mathrm{d}s+\boldsymbol{\Sigma}(\mathbf{Z}_s)\mathrm{d}\mathbf{B}_s\ ,\quad s\in[0,t]\,. \tag{5}$$

In contrast with (1), (5) allows the diffusion coefficient to be state-dependent. Our methods directly apply to all such processes provided there is an explicit transformation $\mathbf{Z}_s\mapsto\boldsymbol{\eta}(\mathbf{Z}_s)=:\mathbf{X}_s$, where $\mathbf{X}$ solves an SDE of the type (1); the implied drift $\boldsymbol{\alpha}$ can be easily expressed in terms of $\boldsymbol{b}$ and $\boldsymbol{\Sigma}$ via Itô's formula and it will have to satisfy the conditions we have already specified. In model (A) the likelihood becomes $f(y_i\mid\boldsymbol{\eta}^{-1}(\mathbf{X}_{t_i}))$, in model (B) the data are $y_i=\zeta(\boldsymbol{\eta}^{-1}(\mathbf{X}_{t_i}))$ and in model (C) the Poisson intensity is $\nu(\boldsymbol{\eta}^{-1}(\mathbf{X}_s))$, where $\boldsymbol{\eta}^{-1}$ denotes the inverse transformation. Therefore, the extension of our methodology to general diffusions is straightforward when $d=1$; under mild conditions (5) can be transformed to (1) by $\eta(Z_s)=\int_{u^*}^{Z_s}\Sigma(z)^{-1}\,\mathrm{d}z$, for some arbitrary u^* in the state space of the diffusion. Moreover, the drift of the transformed process will typically satisfy the three conditions we have specified. However, the extension is harder in higher dimensions. The necessary transformation is more complicated when $d>1$ and it might be intractable or even impossible (Aït-Sahalia, 2004). Even when such a transformation is explicit it might imply a drift for $\mathbf{X}$ which violates condition ii). Nevertheless, many physical systems can be successfully modeled with diffusions which can be transformed to (1).

Our particle filtering methods will be illustrated on two sets of simulated data:

Example 1: Sine diffusion observed with error. The signal satisfies

$$\mathrm{d}X_s = \sin(X_s)\mathrm{d}s + \mathrm{d}B_s\,, \tag{6}$$

and the data consist of noisy observations, $y_i \sim \mathrm{N}(X_{t_i}, \sigma^2)$. Figure 1(top) shows a simulation of this model with $\sigma = 0.2$. In this case

$$\phi(u) = (\sin(u)^2 + \cos(u) + 1)/2\,. \tag{7}$$

This process is closely related to Brownian motion on a circle. It is convenient as an illustrative example since discrete-time skeletons can be easily simulated from this process using the most basic form of the Exact Algorithm (EA1 in Beskos et al., 2006a, R-code is available on request by the authors).

Example 2: OU-driven Cox Process. The second data set consists of the arrival times of a Poisson process, $y_i = t_i$, whose intensity is given by $\nu(X_s), s \geq 0$, where

$$\nu(x) = a + \beta|x|,$$

and X is an Ornstein-Uhlenbeck (OU) process,

$$\mathrm{d}X_s = -\rho X_s \mathrm{d}s + \mathrm{d}B_s\,.$$

The OU process is stationary with Gaussian marginal distribution, $\mathrm{N}(0, 1/(2\rho))$. Thus, an interpretation for this model is that the excursions of X increase the Poisson intensity, whereas a corresponds to the intensity when X is at its mean level. An example data set is shown in Figure 3; where we have taken $a = 0$, $\beta = 20$, $\rho = 1/2$. Although the transition density of the OU process is well-known,

$$X_t \mid X_0 = x_0 \sim \mathrm{N}\left(e^{-\rho t}x_0, \frac{1}{2\rho}(1 - e^{-2\rho t})\right)\,,$$

the observation density $f(y_{i+1} \mid x_{t_i}, x_{t_{i+1}})$ is intractable.

Examples 1 and 2 are examples of observation regimes (A) and (C) respectively. We will show that observation regime (B) can be handled in a similar fashion as (A), so we have not included an accompanying example.

3 Random weight particle filter

As in Section 2 we will denote the observation at time t_i by y_i, and $p_t(\cdot \mid \cdot)$ will denote the system transition density over time t (see Equation 2). We will write $\Delta_i = t_{i+1} - t_i$, and the filtering densities $p(\mathbf{x}_{t_i}|y_{1:i})$ will be denoted by $\pi_i(\mathbf{x}_{t_i})$, where by standard convention $y_{1:i} = (y_1, \ldots, y_i)$. To simplify notation, when we introduce weighted particles below, we will subscript both particles and weights by i rather than t_i.

Our aim is to recursively calculate the filtering densities $\pi_i(\mathbf{x}_{t_i})$. Basic probability calculations yield the following standard filtering recursion for these densities

$$\pi_{i+1}(\mathbf{x}_{t_{i+1}}) \propto \int f(y_{i+1}|\mathbf{x}_{t_i}, \mathbf{x}_{t_{i+1}}) p_{\Delta_i}(\mathbf{x}_{t_{i+1}}|\mathbf{x}_{t_i}) \pi_i(\mathbf{x}_{t_i}) \mathrm{d}\mathbf{x}_{t_i}. \tag{8}$$

Particle filters approximate $\pi_i(\mathbf{x}_{t_i})$ by a discrete distribution, denoted by $\hat{\pi}_i(\mathbf{x}_{t_i})$, whose support is a set of N particles, $\{\mathbf{x}_i^{(j)}\}_{j=1}^N$, with associated probability weight $\{w_i^{(j)}\}_{j=1}^N$. Substituting $\hat{\pi}_i(\mathbf{x}_{t_i})$ for $\pi_i(\mathbf{x}_{t_i})$ in (8), yields a (continuous density) approximation to $\pi_{i+1}(\mathbf{x}_{t_{i+1}})$,

$$\tilde{\pi}_{i+1}(\mathbf{x}_{t_{i+1}}) \propto \sum_{j=1}^{N} w_i^{(j)} f(y_{i+1}|\mathbf{x}_i^{(j)}, \mathbf{x}_{t_{i+1}}) p_{\Delta_i}(\mathbf{x}_{t_{i+1}}|\mathbf{x}_i^{(j)}). \tag{9}$$

The aim of one iteration of the particle filter algorithm is to construct a further particle (discrete distribution) approximation to $\tilde{\pi}_{i+1}(\mathbf{x}_{t_{i+1}})$.

We can obtain such a particle approximation via importance sampling, and a general framework for achieving this is given by the auxiliary particle filter of Pitt and Shephard (1999). We choose a proposal density of the form

$$\sum_{j=1}^{N} \beta_i^{(j)} q(\mathbf{x}_{t_{i+1}}|\mathbf{x}_i^{(j)}, y_{t_{i+1}}) \,. \tag{10}$$

Choice of suitable proposals, i.e. choice of the $\beta_i^{(j)}$s and q, is discussed in the analysis of our specific applications in Section 5.

To simulate a new particle at time t_{i+1} we (a) simulate a particle $\mathbf{x}_i^{(k)}$ at time i, where k is a realisation of a discrete random variable which takes the value $j \in \{1, 2, \ldots, N\}$ with probability $\beta_i^{(j)}$; and (b) simulate a new particle at time t_{i+1} from $q(\mathbf{x}_{t_{i+1}}|\mathbf{x}_i^{(k)}, y_{i+1})$. The weight assigned to this pair of particles $(\mathbf{x}_i^{(k)}, \mathbf{x}_{t_{i+1}})$ is proportional to

$$\frac{w_i^{(k)} f(y_{i+1}|\mathbf{x}_i^{(k)}, \mathbf{x}_{t_{i+1}}) p_{\Delta_i}(\mathbf{x}_{t_{i+1}}|\mathbf{x}_i^{(k)})}{\beta_i^{(k)} q(\mathbf{x}_{t_{i+1}}|\mathbf{x}_i^{(k)}, y_{i+1})} \,. \tag{11}$$

This is repeated N times to produce the set of weighted particles at time t_{i+1}, $\left\{(\mathbf{x}_{i+1}^{(j)}, w_{i+1}^{(j)})\right\}_{j=1}^{N}$, which gives an importance sampling approximation to $\pi_{i+1}(\mathbf{x}_{t_{i+1}})$. Renormalising the weights is possible but does not materially affect the methodology or its accuracy. Improvements on independent sampling in step (a) can be made: see the stratified sampling ideas of Carpenter et al. (1999). The resulting particle filter has good theoretical properties including consistency (Crisan, 2001) and central limit theorems for estimates of posterior moments (Del Moral and Miclo, 2000; Chopin, 2004; Künsch, 2005), as $N \to \infty$. Under conditions relating to exponential forgetting of initial conditions, particle filter errors stabilise as $n \to \infty$ (Del Moral and Guionnet, 2001; Künsch, 2005).

The difficulty with implementing such a particle filter when the signal $\mathbf{X}$ is a diffusion process is that the transition density $p_{\Delta_i}(\mathbf{x}_{t_{i+1}}|\mathbf{x}_i^{(k)})$ which appears in (11) is intractable for most diffusions of interest, due to the expectation term in (2). Furthermore, for observation model (C) (but also for more general models), the likelihood term $f(y_{i+1}|\mathbf{x}_i^{(k)}, \mathbf{x}_{t_{i+1}})$ given in (4) cannot be calculated analytically.

We circumvent these problems by assigning each new particle a random weight which is a realisation of a random variable whose mean is (11). The construction and simulation of this random variable is developed in Section 4, and it is based on the particular expression for the transition density in (2). The replacement of the weights by positive unbiased estimators is an interesting possibility in more general contexts than the one considered in this paper. Indeed, in Section 3.2 we show that this approach amounts to a convenient augmentation of the state with auxiliary variables.

3.1 Simulation of weights

In all models the weight associated with the pair $(\mathbf{x}_i^{(k)}, \mathbf{x}_{t_{i+1}})$ equals

$$h_{i+1}(\mathbf{x}_i^{(k)}, \mathbf{x}_{t_{i+1}}, y_{i+1})\mu_g(\mathbf{x}_i^{(k)}, \mathbf{x}_{t_{i+1}}, t_i, t_{i+1}) \tag{12}$$

where h_{i+1} is a known function, and for $0 < u < t$,

$$\mu_g(\mathbf{x}, \mathbf{z}, u, t) := \mathbb{E}\left[\exp\left\{-\int_u^t g(\mathbf{W}_s)\mathrm{d}s\right\}\right],$$

where the expectation is taken w.r.t. a d-dimensional Brownian bridge $\mathbf{W}$, starting at time u from $\mathbf{W}_u = \mathbf{x}$ and finishing at time t at $\mathbf{W}_t = \mathbf{z}$.

Models (A) and (B) : For these model types

$$h_{i+1}(\mathbf{x}_i^{(k)}, \mathbf{x}_{t_{i+1}}, y_{i+1}) = \frac{w_i^{(k)} f(y_{i+1}|\mathbf{x}_{t_{i+1}})\mathcal{N}_{\Delta_i}(\mathbf{x}_{t_{i+1}} - \mathbf{x}_i^{(k)}) \exp\{A(\mathbf{x}_{t_{i+1}}) - A(\mathbf{x}_i^{(k)})\}}{\beta_i^{(k)} q(\mathbf{x}_{t_{i+1}}|\mathbf{x}_i^{(k)}, y_{i+1})},$$

and $g = \phi$. In model type (B) the proposal distribution $q(\mathbf{x}_{t_{i+1}}|\mathbf{x}_i^{(k)}, y_{i+1})$ should be chosen to propose only values of $\mathbf{x}_{t_{i+1}}$ such that $\zeta(\mathbf{x}_{t_{i+1}}) = y_{i+1}$; then $f(y_{i+1}|\mathbf{x}_{t_{i+1}}) = 1$.

Model (C): A synthesis of (2), (4) and (11), with $g = \phi + \nu$ gives

$$h_{i+1}(\mathbf{x}_i^{(k)}, \mathbf{x}_{t_{i+1}}, y_{i+1}) = \frac{w_i^{(k)} \nu(\mathbf{x}_{t_{i+1}})\mathcal{N}_{\Delta_i}(\mathbf{x}_{t_{i+1}} - \mathbf{x}_i^{(k)}) \exp\{A(\mathbf{x}_{t_{i+1}}) - A(\mathbf{x}_i^{(k)})\}}{\beta_i^{(k)} q(\mathbf{x}_{t_{i+1}}|\mathbf{x}_i^{(k)}, y_{i+1})},$$

Section 4 shows how to construct for each pair of $(\mathbf{x}, \mathbf{z})$ and times (u, t), with $u < t$, additional *auxiliary* variables $\mathbf{V}$, and a function $r(\mathbf{v}, \mathbf{x}, \mathbf{z}, u, t) \geq 0$, with the property that

$\mathbb{E}[r(\mathbf{V}, \mathbf{x}, \mathbf{z}, u, t) \mid \mathbf{x}, \mathbf{z}] = \mu_g(\mathbf{x}, \mathbf{z}, u, t)$. The auxiliary variables are simulated according to an appropriate conditional distribution $Q_g(\cdot \mid \mathbf{x}, \mathbf{z}, u, t)$, and r is easy to evaluate. Our method replaces in the weight the intractable term μ_g with its unbiased estimator r.

Random Weight Particle Filter (RWPF)

PF0 Simulate a sample $\mathbf{x}_0^{(1)}, \ldots, \mathbf{x}_0^{(N)}$ *from* $p(\mathbf{x}_0)$*, and set* $w_0^{(j)} = 1/N$.
For $i = 0, \ldots, n-1$*, for* $j = 1, \ldots, N$*:*

PF1 calculate the effective sample size of the $\{\beta_i^{(k)}\}$*,* $ESS = (\sum_{k=1}^N (\beta_i^{(k)})^2)^{-1}$*; if* $ESS < C$*, for some fixed constant* C*, simulate* $k_{i,j}$ *from* $p(k) = \beta_i^{(k)}$*,* $k = 1, \ldots, N$ *and set* $\delta_{i+1}^{(j)} = 1$*; otherwise set* $k_{i,j} = j$ *and* $\delta_{i+1}^{(j)} = \beta_i^{(j)}$*;*

PF2 simulate $\mathbf{x}_{i+1}^{(j)}$ *from* $q(\mathbf{x}_{t_{i+1}} | \mathbf{x}_i^{(k_{i,j})}, y_{i+1})$*;*

PF3 simulate $\mathbf{v}_{i+1} \sim Q_g(\cdot \mid \mathbf{x}_i^{(k_{i,j})}, \mathbf{x}_{t_{i+1}}, t_i, t_{i+1})$*;*

PF4 assign particle $\mathbf{x}_{i+1}^{(j)}$ *a weight*

$$w_{i+1}^{(j)} = \delta_{i+1}^{(j)} h_{i+1}(\mathbf{x}_i^{(k_{i,j})}, \mathbf{x}_{i+1}^{(j)}, y_{i+1}) r(\mathbf{v}_{i+1}, \mathbf{x}_i^{(k_{i,j})}, \mathbf{x}_{t_{i+1}}, t_i, t_{i+1}) \,. \tag{13}$$

Notice that this algorithm contains a decision as to whether or not resample particles before propagation in step PF1, with decision being based on the ESS of the $\beta_i^{(j)}$. The constant C can be interpreted as the minimum acceptable effective sample size. (See Liu and Chen 1998 for the rationale of basing resampling on such a condition.) Whether or not resampling occurs will affect the weight given to the new sets of particles, and this is accounted for by different values of $\delta_{i+1}^{(j)}$ in PF1. Optimally, the resampling for step PF1 will incorporate dependence across the N samples; for example the stratified sampling scheme of Carpenter et al. (1999) or the residual sampling of Liu and Chen (1998).

3.2 An equivalent formulation via an augmentation of the state

In the previous section we described a generic sequential Monte Carlo scheme where the exact weights in the importance sampling approximation of the filtering distributions are replaced by positive unbiased estimators. We now show that this scheme is equivalent to applying an ordinary auxiliary particle filter to a model with richer latent structure. We demonstrate this equivalent representation for model types (A) and (B), since an obvious modification of the argument establishes the equivalence for model type (C).

According to our construction, conditionally on $\mathbf{X}_{t_i}$, $\mathbf{X}_{t_{i+1}}$, t_i and t_{i+1}, $\mathbf{V}_{i+1}$ is independent of $\mathbf{V}_j$ and $\mathbf{X}_{t_j}$ for any j different from $i, i+1$. Additionally, it follows easily from the unbiasedness and positivity of r that, conditionally on $\mathbf{X}_{t_i} = \mathbf{x}$, $r(\mathbf{v}_{i+1}, \mathbf{x}, \mathbf{x}_{t_{i+1}}, t_i, t_{i+1})$

is a probability density function for $(\mathbf{X}_{t_{i+1}}, \mathbf{V}_{i+1})$ with respect to the product measure $Leb(\mathrm{d}\mathbf{z}) \times Q_g(\mathrm{d}\mathbf{v} \mid \mathbf{x}, \mathbf{z}, t_i, t_{i+1})$, where Leb denotes the Lebesgue measure.

Consider now an alternative discrete-time model with unobserved states $(\mathbf{Z}_i, \mathbf{V}_i), i = 1, \ldots, n$, $\mathbf{Z}_i \in \mathbf{R}^d$, with a non-homogeneous Markov transition density

$$p_{i+1}(\mathbf{z}_{i+1}, \mathbf{v}_{i+1} \mid \mathbf{z}_i, \mathbf{v}_i) = r(\mathbf{v}_{i+1}, \mathbf{z}_i, \mathbf{z}_{i+1}, t_i, t_{i+1}) \,,$$

(this density is with respect to $Leb \times Q_g$) and observed data y_i with observation density $f(y_{i+1} \mid \mathbf{z}_i, \mathbf{z}_{i+1})$. By construction the marginal filtering distributions of $\mathbf{Z}_i$ in this model are precisely $\pi_i(\mathbf{x}_{t_i})$, i.e. the filtering densities in (8). Consider an auxiliary particle filter applied to this model where we choose with probability $\beta_i^{(j)}$ each of the existing particles $(\mathbf{z}_i^{(j)}, \mathbf{v}_i^{(j)})$, and generate new particles according to the following proposal

$$(\mathbf{z}_{i+1}, \mathbf{v}_{i+1}) \sim q(\mathbf{z}_{i+1} \mid \mathbf{z}_i^{(k)}, y_{i+1}) Q_g(\mathrm{d}\mathbf{v}_{i+1} \mid \mathbf{z}_i^{(k)}, \mathbf{z}_{i+1}, t_i, t_{i+1}) Leb(\mathrm{d}\mathbf{z}_{i+1}) \,,$$

where q is the same proposal density as in (10). The weights associated with each particle in this discrete-time model are tractable and are given by (13). Therefore, the weighted sample $\left\{(\mathbf{z}_{i+1}^{(j)}, w_{i+1}^{(j)})\right\}_{j=1}^{N}$ is precisely a particle approximation to $\pi_{i+1}(\mathbf{x}_{t_{i+1}})$, and RWPF is equivalent to an auxiliary particle filter on this discrete-time model whose latent structure has been augmented with the auxiliary variables $\mathbf{V}_i$.

This equivalent representation sheds light on many aspects of our method. Firstly, it makes it obvious that it is inefficient to average more than one realization of the positive unbiased estimator of μ_g per particle. Instead it is more efficient to generate more particles with only one realization of the estimator simulated for each pair of particles.

Secondly, it illustrates that RWPF combines the advantages of the bootstrap and the auxiliary particle filter. Although it is easy to simulate from the probability distribution Q_g (as described in Section 4), it is very difficult to derive its density. (with respect to an appropriate reference measure). Since the $\mathbf{V}_i$s are propagated according to this measure, its calculation is avoided. This is an appealing feature of the bootstrap filter which propagates particles without requiring analytically the system transition density. On the other hand the propagation of the $\mathbf{Z}_i$s is done via a user-specified density which incorporates the information in the data.

Thirdly, it suggests that the RWPF will have similar theoretical properties with auxiliary particle filters applied to discrete-time models. This is explored in Section 3.3.

3.3 Theoretical properties

Consider estimation of the posterior mean of some function φ of the state at time t_i, $\mathbb{E}[\varphi(\mathbf{x}_{t_i})|y_{1:i}]$. A natural approach to the investigation of particle filter effectiveness is to consider the limiting behaviour of the algorithm as $N \to \infty$. For the standard auxiliary

particle filter, Chopin (2004) introduces a central limit theorem (CLT) for estimation of this type of expectations. This CLT applies directly to both EPPF and the ESPF.

In Appendix F we extend the result of Chopin (2004) and give a further necessary condition on the random weights in RWPF under which a CLT still holds. This extra condition is (C2). The expression for the variance of the estimator of $\mathbb{E}[\varphi(\mathbf{x}_{t_i})|y_{1:i}]$ obtained with RWPF differs from the expression in the standard case (i.e. when the weights are known) by an extra term caused by the randomness in the weights (see Equations 27–29 and the comment on Theorem 3 in Appendix F for further details). The ready adaptation of Chopin's approach is facilitated by the observation that the RWPF can be re-expressed as a standard particle filter for the an augmented state (see Section 3.2).

One important consequence of this CLT is that the errors in estimating $\mathbb{E}[\varphi(\mathbf{x}_{t_i})|y_{1:i}]$ are of order $N^{-1/2}$. Previous filtering methods for the diffusion problems we consider are based on (i) discretising time and introducing M intermediate time points between each observation time; (ii) using an Euler, or higher order, approximation to the diffusion (1); and (iii) applying a particle, or other, filter to this approximate discrete time model. See for example Crisan et al. (1999). Results giving the order of the errors in one such scheme are given by Del Moral et al. (2001). For Models such as (A) and (B) the errors are of order $N^{-1/2}$ provided that the number of intermediate time steps M between each observation increases at a rate $N^{1/2}$. Thus for fixed computational cost $K \propto MN$ the errors decrease at a rate $K^{-1/3}$. For models such as (C), where the likelihood depends on the path of the state between two successive observations, the rate at which errors decrease will be slower, for example $K^{-1/4}$ (Del Moral et al., 2001), or $K^{-1/6}$ (Theorem 1.1 of Crisan et al., 1999).

4 Generalised Poisson Estimators

We have already motivated the need for the simulation of a positive unbiased estimator of

$$\mathbb{E}\left[E\right] \text{ where } E =: \exp\left\{-\int_0^t g(\mathbf{W}_s)\mathrm{d}s\right\}, \tag{14}$$

where the expectation is taken w.r.t. a d-dimensional Brownian bridge $\mathbf{W}$. In this section we introduce a methodology for deriving such estimators, and provide theoretical and simulation results regarding the variance of the suggested estimators. These results are of independent interest beyond particle filtering, so we present our methodology in a general way, where g is an arbitrary function assumed only to be continuous on $\mathbf{R}^d$. We assume that $\mathbf{W}_0 = \mathbf{x}$ and $\mathbf{W}_t = \mathbf{z}$, for arbitrary $\mathbf{x}, \mathbf{z} \in \mathbf{R}^d$ and $t > 0$. By the time-homogeneity property of the Brownian bridge our methodology extends to the case where the integration limits change to u and $u + t$, for any $u > 0$.

Beskos et al. (2006b) proposed an unbiased estimator of (14), the *Poisson Estimator*:

$$\text{PE:} \quad e^{(\lambda-c)t}\lambda^{-\kappa}\prod_{j=1}^{\kappa}\left[c - g(\mathbf{W}_{\psi_j})\right]; \tag{15}$$

κ is a Poisson random variable with mean λt, the ψ_js are uniformly distributed on $[0,t]$, and $c \in \mathbf{R}$, $\lambda > 0$ are arbitrary constants. (Here and below we assume that the empty product, i.e. when $\kappa = 0$, takes the value 1.) The two main weaknesses of the PE are that it may return negative estimates and that its variance is not guaranteed to be finite. Both of these problems are alleviated when g is bounded. However this is a very restrictive assumption in our context. Therefore, here, we introduce a collection of unbiased and positive estimators of (14) which generalise the PE. The methods we consider allow c and λ depend on $\mathbf{W}$, and permit κ to have a general discrete distribution. Firstly, we need to be able to simulate random variables $L_{\mathbf{W}}$ and $U_{\mathbf{W}}$ with

$$L_{\mathbf{W}} \leq g(\mathbf{W}_s) \leq U_{\mathbf{W}}, \quad \text{for all } s \in [0,t], \tag{16}$$

and to be able to simulate $\mathbf{W}_s$ at any s, given the condition implied by (16). For unbounded g this is non-trivial. However, both of these simulations have become feasible since the introduction of an efficient algorithm in Beskos et al. (2005b). An outline of the construction is given in Appendix A.

Let $U_{\mathbf{W}}$ and $L_{\mathbf{W}}$ satisfy (16) and $\psi_j, j \geq 1$, be a sequence of independent uniform random variables on $[0,t]$. Then, (14) can be re-expressed as follows,

$$\begin{aligned}
\mathbb{E}\left[e^{-U_{\mathbf{W}}t}\exp\left\{\int_0^t (U_{\mathbf{W}} - g(\mathbf{W}_s))\mathrm{d}s\right\}\right] &= \mathbb{E}\left[e^{-U_{\mathbf{W}}t}\sum_{k=0}^{\infty}\frac{1}{k!}\left(\int_0^t (U_{\mathbf{W}} - g(\mathbf{W}_s))\mathrm{d}s\right)^k\right] \\
&= \mathbb{E}\left[e^{-U_{\mathbf{W}}t}\mathbb{E}\left[\sum_{k=0}^{\infty}\frac{t^k}{k!}\prod_{j=1}^{k}(U_{\mathbf{W}} - g(\mathbf{W}_{\psi_j})) \mid U_{\mathbf{W}}, L_{\mathbf{W}}\right]\right] \\
&= \mathbb{E}\left[e^{-U_{\mathbf{W}}t}\frac{t^\kappa}{\kappa! p(\kappa \mid U_{\mathbf{W}}, L_{\mathbf{W}})}\prod_{j=1}^{\kappa}(U_{\mathbf{W}} - g(\mathbf{W}_{\psi_j}))\right],
\end{aligned} \tag{17}$$

where κ is a discrete random variable with conditional probabilities $\mathrm{P}[\kappa = k \mid U_{\mathbf{W}}, L_{\mathbf{W}}] = p(k \mid U_{\mathbf{W}}, L_{\mathbf{W}})$. The second equality in the above argument is obtained using dominated convergence and Fubini's theorem (which hold by positivity of the summands).

We can derive various estimators of (14) by specifying $p(\cdot \mid U_{\mathbf{W}}, L_{\mathbf{W}})$. The family of all such estimators will be called the *Generalised Poisson Estimator* (GPE):

$$\text{GPE:} \quad e^{-U_{\mathbf{W}}t}\frac{t^\kappa}{\kappa! p(\kappa \mid U_{\mathbf{W}}, L_{\mathbf{W}})}\prod_{j=1}^{\kappa}(U_{\mathbf{W}} - g(\mathbf{W}_{\psi_j}))\,. \tag{18}$$

The following Theorem (proved in Appendix B) gives the optimal choice for $p(\cdot \mid U_{\mathbf{W}}, L_{\mathbf{W}})$.

Theorem 1. *The conditional second moment of the Generalised Poisson Estimator given* $U_{\mathbf{W}}$ *and* $L_{\mathbf{W}}$*, is:*

$$e^{-2U_{\mathbf{W}}t}\sum_{k=0}^{\infty}\frac{t^k}{p(k\mid U_{\mathbf{W}},L_{\mathbf{W}})k!^2}\mathbb{E}\left[\left(\int_0^t(U_{\mathbf{W}}-g(\mathbf{W}_s))^2ds\right)^k\mid U_{\mathbf{W}},L_{\mathbf{W}}\right]. \tag{19}$$

If

$$\sum_{k=0}^{\infty}\frac{t^{k/2}}{k!}\mathbb{E}\left[\left(\int_0^t(U_{\mathbf{W}}-g(\mathbf{W}_s))^2ds\right)^k\mid U_{\mathbf{W}},L_{\mathbf{W}}\right]^{1/2}<\infty\,, \tag{20}$$

then the second moment is minimised by the choice

$$p(k\mid U_{\mathbf{W}},L_{\mathbf{W}})\propto\frac{t^{k/2}}{k!}\mathbb{E}\left[\left(\int_0^t(U_{\mathbf{W}}-g(\mathbf{W}_s))^2ds\right)^k\mid U_{\mathbf{W}},L_{\mathbf{W}}\right]^{1/2}, \tag{21}$$

with minimum second moment given by

$$\left(e^{-U_{\mathbf{W}}t}\sum_{k=0}^{\infty}\frac{t^{k/2}}{k!}\mathbb{E}\left[\left(\int_0^t(U_{\mathbf{W}}-g(\mathbf{W}_s))^2ds\right)^k\mid U_{\mathbf{W}},L_{\mathbf{W}}\right]^{1/2}\right)^2<\infty\,,\ \textit{for almost all}\ U_{\mathbf{W}},L_{\mathbf{W}}\,. \tag{22}$$

Whilst the right-hand side of (21) cannot be evaluated analytically, it can guide a suitable choice of $p(\cdot\mid U_{\mathbf{W}},L_{\mathbf{W}})$. If $\mathbf{W}$ were known, the optimal proposal is Poisson with mean

$$\lambda_{\mathbf{W}}:=\left(t\int_0^t(U_{\mathbf{W}}-g(\mathbf{W}_s))^2\mathrm{d}s\right)^{1/2}. \tag{23}$$

We will discuss two possible ways that (23) can be used to choose a good proposal.

A conservative approach takes $p(\cdot\mid U_{\mathbf{W}},L_{\mathbf{W}})$ to be Poisson with mean $(U_{\mathbf{W}}-L_{\mathbf{W}})t$ (an upper bound of $\lambda_{\mathbf{W}}$). We call this estimator GPE-1. An advantage of GPE-1 is that its second moment is bounded above by $\mathbb{E}[e^{-2L_{\mathbf{W}}t}]$. Thus, under mild and explicit conditions on g, which are contained in the following theorem (proved in Appendix C), the variance of the estimator is guaranteed to be finite.

Theorem 2. *A sufficient condition for GPE-1 to have finite variance is that*

$$g(u_1,\ldots,u_d)\geq-\delta\sum_{i=1}^{d}(1+|u_i|),\quad\textit{for all}\ u_i\in\mathbf{R},1\leq i\leq d,\ \delta\geq0.$$

Since $\lambda_{\mathbf{W}}$ is stochastic, an alternative approach is to introduce a (exogenous) random mean and assume that $p(\cdot|U_{\mathbf{W}}, L_{\mathbf{W}})$ is Poisson with this random mean. For tractability we choose the random mean to have a Gamma distribution, when $p(\cdot \mid U_{\mathbf{W}}, L_{\mathbf{W}})$ becomes a negative-binomial distribution:

$$\text{GPE-2:} \quad e^{-U_{\mathbf{W}}t}\frac{t^{\kappa}\Gamma(\beta)(\beta+\gamma_{\mathbf{W}})^{\beta+\kappa}}{\Gamma(\beta+\kappa)\beta^{\beta}\gamma_{\mathbf{W}}^{\kappa}}\prod_{j=1}^{\kappa}\left[U_{\mathbf{W}} - g(\mathbf{W}_{\psi_j})\right], \tag{24}$$

where $\gamma_{\mathbf{W}}$ and β denote the mean and the dispersion parameter respectively of the negative binomial. Since the negative-binomial has heavier tails than the Poisson Estimator, GPE-2 will have finite variance whenever there exists a PE with finite variance. On the other hand, big efficiency gains can be achieved if $\gamma_{\mathbf{W}}$ is chosen to be approximately $\mathbb{E}[\lambda_{\mathbf{W}} \mid U_{\mathbf{W}}, L_{\mathbf{W}}]$. There is a variety of ad-hoc methods which can provide a rough estimation of this expectation. Applying Jensen's inequality to exchange the integration with the square power in (23), and subsequently approximating $\mathbb{E}[\, g(\mathbf{W}_s) \mid U_{\mathbf{W}}, L_{\mathbf{W}}\,]$ by $g(\mathbb{E}[\mathbf{W}_s])$, suggests taking

$$\gamma_{\mathbf{W}} = tU_{\mathbf{W}} - \int_0^t g\left(\mathbf{x}\frac{t-s}{t} + \mathbf{y}\frac{s}{t}\right)\mathrm{d}s > 0\,. \tag{25}$$

A simulation study (part of which is presented in Section 4.1 below) reveals that this choice works very well in practice and the GPE-2 has up to several orders of magnitude smaller variance than the PE or the GPE-1. The integral can usually be easily evaluated, otherwise a crude approximation can be used.

We have confined our presentation to the case where the expectation in (14) is w.r.t. the Brownian bridge measure. Nevertheless, as pointed out in Beskos et al. (2006b) the PE can be constructed in exactly the same way when the expectation is taken w.r.t. an arbitrary diffusion bridge measure, as long as exact skeletons can be simulated from this measure. The GPE can also be implemented in this wider framework, provided that the process $\mathbf{W}$ can be constructed to satisfy (16).

4.1 Simulation study

We consider a smooth bounded test function $g(u) = (\sin(u)^2 + \cos(u) + 1)/2$. This has been chosen in view of Example 1. The function g is periodic, with period 2π. In $[0, 2\pi]$ it has local minima at 0 and 2π, global minimum at π and maxima at $\pi/3$ and $5\pi/3$. Since g is bounded by $9/8$ we can construct a PE which returns positive estimates by setting $c \geq 9/8$. Under this constraint, Beskos et al. (2006b) argued that a good choice is $c = \lambda = 9/8$. Simulation experiments suggested that the performance of the GPE-2 is quite robust to the choice of the dispersion parameter β. We have fixed it in our examples to $\beta = 10$. Table 1 summarizes estimates of the variance of the estimators based on 10^4 simulated values. We see that GPE-2 can be significantly more efficient than PE, in particular when taking into account $\mathbb{E}[\kappa]$. In general, the performance of PE is sensitive to the choice of c and λ.

	Estimator	$x = 0, z = 0$	$x = 0, z = \pi$	$x = \pi, z = \pi$
variance	PE	0.202	0.200	0.027
	GPE-1	4.21×10^{-3}	0.208	0.034
	GPE-2	2.08×10^{-3}	0.220	0.033
	Var(E)	3.74×10^{-5}	3.27×10^{-3}	4.72×10^{-3}
$\mathbb{E}[\kappa]$	PE	1.118	1.126	1.121
	GPE-1	0.130	1.091	0.744
	GPE-2	0.119	0.329	0.735

Table 1: Monte Carlo estimates of the variance of four estimators of (14) where $g(u) = (\sin(u)^2 + \cos(u) + 1)/2$. For comparison we give also var(E). We also report an estimate of $\mathbb{E}[\kappa]$. We consider three different pairs of starting and ending points (x, z) and time increment $t = 1$. The estimates in the table were obtained from a sample of 10^4 realisations.

GPE-1 is typically less efficient than GPE-2. Table 1 also gives the value of Var(E) which takes significantly smaller values (by a couple of orders of magnitude) than any of PE, GPE-1 or GPE-2, illustrating the efficiency cost of these auxiliary variable constructions in absolute terms.

We have also investigated how the efficiency of the PE and GPE-2 varies with the time increment t and in particular for small t (results not shown). These empirical results suggest that the coefficient of variation of the errors of both PE and GPE-2 are $O(t^\delta)$ for some $\delta > 0$; but that the value of δ differs for the two estimators. In the cases that we investigated, the GPE-2 appears to have a faster rate of convergence than PE.

The results of this simulation study have been verified for other functions g (results not shown). We have experimented with differentiable (e.g. $g(u) = u$) and non-differentiable (e.g. $g(u) = |u|$) unbounded functions. In these cases it is impossible to design a PE which returns positive estimates w.p.1. Again, we have found that the GPE-2 performs significantly better than the PE.

It is important to mention that alternative Monte Carlo methods exist which yield consistent but biased estimates of (14). One such estimator is obtained by replacing the time-integral in (14) with a Riemann approximation based on a number, M say, of intermediate points. This technique is used to construct a transition density estimator in Nicolau (2002) and effectively underlies the transition density estimator of Durham and Gallant (2002) (when the diffusion process has constant diffusion coefficient). The approach of Durham and Gallant (2002) has been used in MCMC and filtering applications (Golightly and Wilkinson, 2006; Chib et al., 2006; Ionides, 2003). In the filtering context it provides an alternative to RWPF, where the weights are approximated. It is not the purpose of this paper to carry out a careful comparison of RWPF with such variants. However, as an illustration we present a very small scale comparison in the context of estimating the transition density, $p_t(z \mid x)$, of (6) for $t = 1$ and x, z as in Table 1. We compare 4 meth-

Estimator	$x = 0, z = 0$	$x = 0, z = \pi$	$x = \pi, z = \pi$
PE	1.25	0.93	0.17
GPE-2	0.13	0.78	0.2
DG-1	0.5	0.45	0.3
DG-5	0.28	0.19	0.22

Table 2: Monte Carlo estimates based on 10^4 realisations of the root mean square error divided by the true value of 4 estimators of $p_t(z \mid x)$, of (6) for $t = 1$ and various x, z. As true value we take the estimate produced by averaging the estimations given by GPE-2. The number of intermediate points used for each estimator are 1 and 5 for DG-1 and DG-5 respectively; the number of Brownian bridge simulations for PE and GPE-2 are given in Table 1 ($\mathbb{E}[\kappa]$).

ods. Two are based on (2) and use the PE and the GPE-2 to generate estimators of the expectation. The other two, DG-1 and DG-5 are two implementation of the Durham and Gallant (2002) estimator, with 1 and 5 respectively intermediate points. We compare the methods in terms of their root mean square error divided by the true value (i.e. the coefficient of variation). As the true value we used the estimate of the GPE-2. The results of the comparison are presented in Table 2. Notice that DG-1 and DG-5 simulate many more variables than GPE-2 to construct their estimates.

5 Comparison of particle filters on the simulated data

We now demonstrate the performance of the different particle filters we have presented on the two examples introduced in Section 2.

5.1 Analysis of the sine diffusion

We first consider analysing the sine diffusion of Example 1. The simulated data is shown in Figure 1(top). We compare four implementations of the particle filter each of which avoids time-discretisations by using methodology based on the Exact Algorithm (EA) for simulating diffusions: i) EPPF, which uses EA for implementing a bootstrap filter, ii) ESPF, which adapts EA to simulate by rejection sampling from the filtering densities, iii) RWPF1, an implementation of RWPF using PE (see Table 1) to simulate the weights, iv) RWPF2, an implementation of RWPF using GPE-2 to simulate the weights. Details on the implementation of EPPF and ESPF are given in Appendix D.

In this simple example ESPF is more efficient than EPPF, since it has the same computational cost, but it is proposing from the optimal proposal distribution. However, we have efficiently implemented ESPF exploiting several niceties of this simple model, in particular the Gaussian likelihood and the fact that the drift is bounded. In more general models

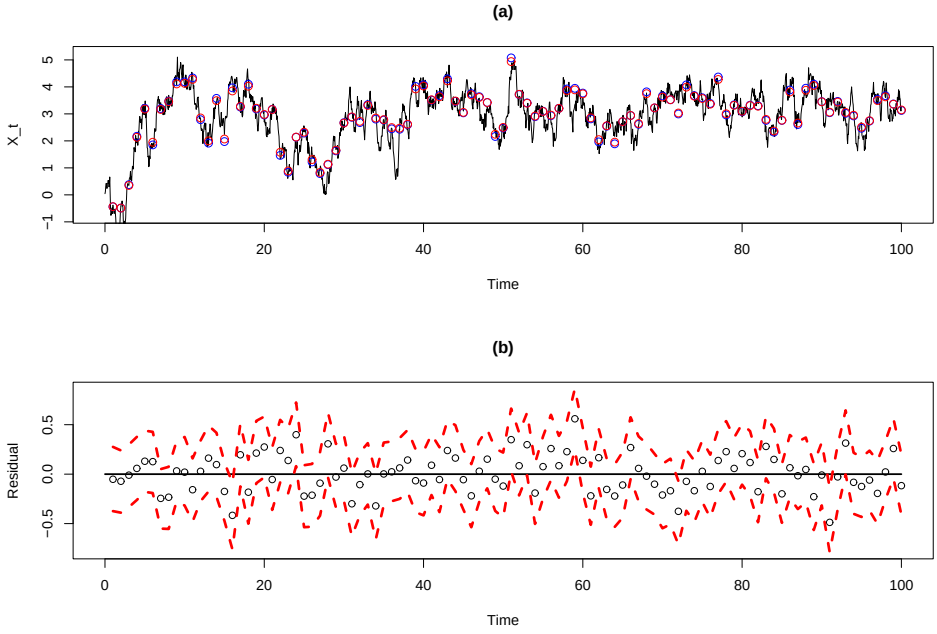

Figure 1: Top: A realisation of the sine diffusion (black line) on $[0, 100]$; 100 observations at unit time intervals (blue circles); mean of the filtering distribution of the diffusion at the observation times obtained by RWPF2 with $N = 1,000$ particles (red circles). Bottom: the difference between observed data and filtered means (circles) and 90% credible intervals (red dashed lines) from RWPF2. (Whilst for clarity they are shown for all times, the credible intervals were only calculated at the observation times.)

implementation of ESPF can be considerably harder and its comparison with EPPF less favorable due to smaller acceptance probabilities.

In this context where ϕ is bounded one can speed up the implementation of GPE-2 with practically no loss of efficiency by replacing $U_{\mathbf{W}}$ in (24) and (25) by 9/8 which is the upper bound of ϕ. In this case, there is no need to simulate $U_{\mathbf{W}}$ and $L_{\mathbf{W}}$. We have implemented this simplification in the RWPF2.

Algorithms EPPF, RWPF1-2 used the stratified re-sampling algorithm of Carpenter et al. (1999), with re-sampling at every iteration. For RWPF1-2 we chose the proposal distribution for the new particles based on the optimal proposal distribution obtained if the sine diffusion is approximated by the Ozaki discretisation scheme (details in Appendix E). For EPPF we chose the $\beta_i^{(k)}$s to be those obtained from this approximation.

The number of particles used in each algorithm was set so that each filter had comparable CPU cost, which resulted in 500, 500, 910 and 1000 particles used respectively for each algorithm. For these numbers of particles, EPPF and ESPF on average required

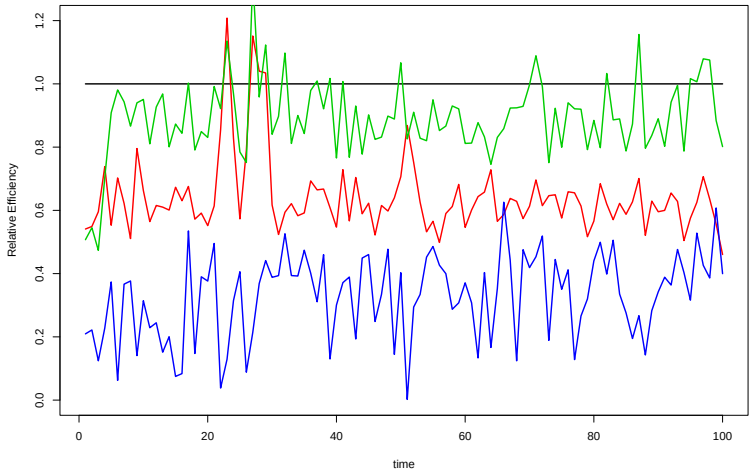

Figure 2: Relative efficiency of the 4 particle filter algorithms at estimating the filtering mean $\mathbb{E}[X_{t_i}|y_{1:i}]$. Each line gives the relative efficiency of one algorithm compared to RWPF2 (black: RWPF2, green: RWPF1, red: ESPF, blue: EPPF). See text for details.

the proposal of 1360 particles and required 675 Brownian bridge simulations within the accept-reject step (iii) at each iteration of the algorithm. By comparison RWPF1 and RWPF2 simulated respectively 910 and 1000 particles and required on average 1025 and 850 Brownian bridge simulations to generate the random weights at each iteration.

Note that the comparative CPU cost of the four algorithms, and in particular that of EPPF and ESPF as compared to RWPF1-2 depends on the underlying diffusion path. The acceptance probabilities within EPPF and ESPF depend on the values of $x_i^{k_{i,j}}$ and $x_{t_{i+1}}$, and get small when both these values are close to $0 (\mathrm{mod}\ 2\pi)$. (in the long run the diffusion will visit these regions infrequently and will stay there for short periods.) Thus, simulated paths which spent more (or less) time in this region of the state-space would result in EPPF and ESPF having a larger (respectively smaller) CPU cost.

We compared the four filters based on the variability of estimates of the mean of the filtering distribution of the state across 500 independent runs of each filter. Results are given in Figure 2, while output from one run of RWPF2 is shown in Figure 1. The comparative results in Figure 2 are for estimating the mean of the filtering distribution at each iteration (similar results were obtained for various quantiles of the filtering distribution). They show RWPF2 performing best with an average efficiency gain of 15% over RWPF1, 50% over ESPF and 200% over EPPF. Interpretation of these results suggest that (for example) ESPF would be required to run with $N = 750$ (taking 1.5 times the CPU cost for this data set) to obtain comparable accuracy with RWPF2.

Varying the parameters of the model and implementation of the algorithms will affect

the relative performance of the algorithms. In particular increasing (or decreasing) σ^2, the variance of the measurement error, will increase (respectively decrease) the relative efficiency of EPPF relative to the other filters. Similar results occur as Δ_i is decreased (respectively increased). The relative performance of the other three algorithms appears to be more robust to such changes. We considered implementing EPPF with $\beta_i^{(k)} = w_i^{(k)}$; and also using an Euler rather than an Ozaki approximation of the sine diffusion to construct the proposal distribution for RWPF1-2, but neither of these changes had any noticeable effect on the performance of the methods. We also considered re-sampling less often, setting $C = N/4$ in step PF1 of the RWPF algorithm (so re-sampling when the effective sample size of the $\beta_i^{(j)}$s was less than $N/4$) and this reduced the performance of the algorithms substantially (by a factor of 2 for RWPF1-2).

We also investigated the effect of increasing the amount of time, Δ, between observations. To do this we used the above data taking (i) every 10th; or (ii) every 20th time-point.

To measure the performance of the filter for these different scenarios we used the Effective Sample Size (ESS) of Carpenter et al. (1999). ESS is calculated based on the variance of estimates of posterior means across independent runs of the filter, but this variance is compared to the posterior variance to give some measure of how many independent draws from the posterior would produce estimators of the same level of accuracy. We focus on estimates of the posterior mean of the state at observation times; and if s^2 is the sample variance of the particle filter's estimate of $\mathbb{E}[X_{t_i}|y_{1:i}]$ across 100 independent runs, and $\hat{\sigma}^2$ is an estimate of $\mathrm{Var}[X_{t_i}|y_{1:i}]$, then the ESS is $\hat{\sigma}^2/s^2$. Note that comparing filters by their ESS is equivalent to comparing filters based on the variance of the estimators.

Table 3 gives ESS values for the different values of Δ. We see that the ESS values drops dramatically as Δ increases, and the filter is inefficient for $\Delta = 20$. This drop in performance is due to the large variability of the random weights in this case. The variability of these weights is due to (a) the variability of

$$\exp\left\{-\int_{t_i}^{t_{i+1}} g(\mathbf{W}_s)\mathrm{d}s\right\}, \tag{26}$$

across different diffusion paths; and (b) the Monte Carlo variability in estimating this for a given path. To evaluate what amount is due to (a), we tried a particle filter that estimates (26) numerically by simulating the Brownian Bridge at a set of discrete time points (for this example we sampled values every 1/2 time unit) and then using these to numerically evaluate the integral. This approach is closely related to the importance sampling approach of Durham and Gallant (2002); Nicolau (2002), see Section 4.1. The results for this filter are also given in Table 3 (note the ESS values ignore any bias introduced through this numerical approximation), and we again see small ESS values, particularly for $\Delta = 20$. This filter's performance is very similar to the RWPF, which suggests that the Monte Carlo variability in (b) is a small contributor to the poor performance of the RWPF in this case.

Finally we tried introducing pseudo observations at all integer time-intervals where currently no observation is made. The RWPF is then run as above, but with no likelihood

Filter	$\Delta = 10$	$\Delta = 20$
RWPF2	73	5
Discretisation	80	12
pseudoRWPF2	923	933

Table 3: Comparison of filter's mean ESS values for different time intervals between observations (Δ). Results are for the Random Weight Particle Filter using GPE-2 (RWPF2), a filter that numerically approximates the weight through discretising the diffusion process (Discretisation), and the RWPF after introducing uninformative observations at unit time intervals (pseudoRWPF2).

contribution to the weight at the time-points where there are these uninformative observations. The idea is that now $\Delta = 1$, so that the variance of the random weights is well-behaved, but we still have adaptation of the path of the diffusion in the unit time-interval prior to an observation to take account of the information in that observation. Results are again shown in Table 3, and the ESS values are very high (and close to the optimal value, that of the number of particles, 1000). Note that the computational cost is only roughly doubled by adding these extra pseudo observations; as the total computational cost for the simulation of the Brownian bridge is unchanged. These results are reasonably robust to the choice of how frequently to introduce these uninformative observations (results not shown).

5.2 Analysis of the Cox process

We now consider applying the random weight particle filter (RWPF) to Example 2 from Section 2, the OU-driven Cox process. The data we analysed is given in Figure 3(top). It is either impossible or difficult to adapt the other two EA-based particle filters (the EPPF and the ASPF) to this problem. For instance we cannot implement EPPF as the likelihood function is not tractable. As such we just focus on the efficiency of the RWPF in estimating the filtering distribution of $|X_t|$.

Our implementation of the RWPF was based on proposing particles from the prior distribution, so $\beta_i^{(k)} = w_i^{(k)}$ and $q(x_{t_{i+1}}|x_i^j, y_{i+1})$ is just the OU transition density $p(x_{t_{i+1}}|x_i^j)$. We simulated the random weights by GPE-2. We calculated the filtering density at each observation time, and also at 56 pseudo-observation times chosen so that the maximum time difference between two consecutive times for which we calculated the filtering density was 0.1. This was necessary to avoid the number of Brownian bridge simulations required to simulate the weights being too large for long inter-observation times, and also to control the variance of the random weights (see above). The likelihood function for these non-observation times is obtained by removing $\nu(x_{t_i})$ from (4).

We set the number of particles to $1,000$ and resampled when the ESS of the $\beta_i^{(j)}$s was

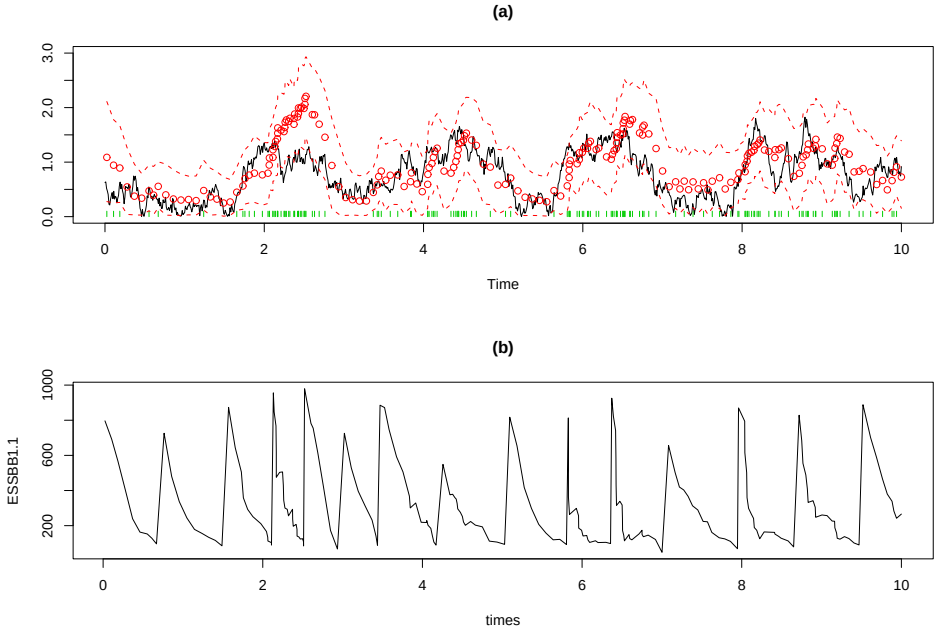

Figure 3: Top: Simulation from the Cox process of Example 2 and results from analysis by the RWPF. The path of the absolute of the underlying diffusion (black line); observed arrival times (green dashes); filter estimates from the RWPF (red circles); and 90% credible interval for the absolute of the diffusion (red dashed line). (Whilst for clarity they are shown for all time, the credible intervals were only calculated at and apply for times where filter estimates are shown.) Bottom: ESS of the RWPFs weights (defined as $(\sum_{j=1}^{N} w_i^{(j)})^2 / \sum_{j=1}^{N} (w_i^{(j)})^2$) over time. The dramatic increases in the effective sample sizes correspond to re-sampling times.

less then 100 ($C = N/10$ in step PF1 of the algorithm in Section 3). Whilst results for the sine diffusion suggest that this will result in an algorithm that re-samples too infrequently, we chose to have a low threshold so that we could monitor the performance of the particle filter by how the ESS of the particle filter weights decay over time. The results of one run of this filter are shown in Figure 3(top). The computational efficiency of this method can be gauged by Figure 3 (bottom) where the ESS of the $w_i^{(j)}$s is plotted over time.

6 Discussion

We have described how recent methods for the exact simulation of diffusions and the unbiased estimation of diffusion exponential functionals can be used within particle filters, so that the resulting particle filters avoid the need for time-discretisation. Among the

approaches we have introduced special attention was given to RWPF which implements an auxiliary particle filter, but simulates the weights that are allocated to each particle. We showed that this methodology is equivalent to an auxiliary particle filter applied to appropriately expanded model. We expect that this methodology will have interesting applications to different models than those considered in this paper, which however involve intractable dynamics or likelihoods.

We have focused on the filtering problem, estimating the current state given observations to date. However, extensions to prediction are trivial – merely requiring the ability to simulate from the state equation, which is possible via the EA algorithms. It is also straightforward to use the idea of Kitagawa (1996), where each particle stores the history of its trajectory, to get approximations of the smoothing density (the density of the state at some time in the past given the observations to date).

Note that while particles store values of the state only for each observation time, it is straightforward to fill in the diffusion paths between these times to produce inferences about the state at any time. A particle approximation to the distribution of $(\mathbf{X}_s, t_{i-1} < s < t_i)$, conditionally on the data $y_{1:i}$ can be constructed using the current set of weighted particles $\{(\mathbf{x}_{i-1}^{(j)}, \mathbf{x}_i^{(j)})\}_{j=1}^N$ with weights $\{w_i^{(j)}\}$, as follows. Firstly we need to introduce some notation; we denote by $\mathbf{x}_{i-1|i}^{(j)}$ the value of the particle at time t_{i-1} from which the jth particle at time t_i is descended. The particle approximation is given by a set of weighted paths $\{(\mathbf{x}_s, t_{i-1} < s < t_i)^{(j)}\}_{j=1}^N$ with weights $\{w_i^{(j)}\}$. Each path is a diffusion bridge starting from $\mathbf{x}_{i-1|i}^{(j)}$ and finishing at $\mathbf{x}_i^{(j)}$ and it can be simulated using EA, as described in Beskos et al. (2006a) and Beskos et al. (2005b). In observation regimes (A) and (B) the EA is applied to simulate a diffusion bridge with density w.r.t. the Brownian bridge measure given by $\exp\{-\int_{t_{i-1}}^{t_i} \phi(\mathbf{X}_s)\mathrm{d}s\}$, whereas in regime (C) the corresponding density is $\exp\{-\int_{t_{i-1}}^{t_i} (\phi(\mathbf{X}_s) + \nu(\mathbf{X}_s))\mathrm{d}s\}$. This representation can be directly exploited to draw inferences for any function of a finite skeleton of $\mathbf{X}$ in-between observation times.

Appendix A: The layered Brownian motion

The algorithm proposed in Beskos et al. (2005b) starts by creating a partition of the sample space of $\mathbf{W}$ for the given $\mathbf{W}_0 = \mathbf{x}$ and $\mathbf{W}_t = \mathbf{y}$. Writing $\mathbf{x} = (x_1, \ldots, x_d)$, for a user-specified constant $a > \sqrt{t/3}$, a sequence of subsets of $\mathbf{R}^d$ is formed as $A_j = \{\mathbf{u} = (u_1, \ldots, u_d) : \min(x_{t_i}, y_i) - ja < u_i \leq \max(x_{t_i}, y_i) + ja\},\ j \geq 0$, where $\cup_j A_j = \mathbf{R}^d$. This sequence defines a partition of the sample space of the form $\cup_{j=1}^{\infty} D_j$, where a path belongs to D_j if and only if the path has exceeded the bounds determined by A_{j-1} but not the bounds determined by A_j. In Beskos et al. (2005b) it is shown how to simulate the random variable which determines which of the D_js $\mathbf{W}$ belongs to, and how to simulate $\mathbf{W}$ at any collection of times conditional on this random variable, the *layered Brownian bridge* construction. Since g is assumed continuous, knowing $W \in D_j$ is sufficient to determine $U_{\mathbf{W}}$ and $L_{\mathbf{W}}$ which satisfy (16). In fact, in the simplified setting where g is bounded, as in

the sine diffusion of Example 1, the layered Brownian bridge construction can be avoided since it is easy to choose $U_{\mathbf{W}}$ and $L_{\mathbf{W}}$ independently of $\mathbf{W}$.

Appendix B: Proof of Theorem 1

$$I := \frac{t^\kappa}{\kappa! p(\kappa \mid U_{\mathbf{W}}, L_{\mathbf{W}})} \prod_{j=1}^{\kappa} (U_{\mathbf{W}} - g(\mathbf{W}_{\psi_j})) \,.$$

Then, (19) is established as follows:

$$\begin{aligned}
\mathbb{E}[I^2 \mid U_{\mathbf{W}}, L_{\mathbf{W}}] &= \mathbb{E}[\,\mathbb{E}[I^2 \mid \kappa, \mathbf{W}]\,] = \mathbb{E}\left[\frac{t^{2\kappa}}{(\kappa! p(\kappa \mid U_{\mathbf{W}}, L_{\mathbf{W}}))^2} \left(\int_0^t \frac{(U_{\mathbf{W}} - g(\mathbf{W}_s))^2}{t} \mathrm{d}s\right)^\kappa\right] \\
&= \mathbb{E}\left[\frac{t^\kappa}{(\kappa! p(k \mid U_{\mathbf{W}}, L_{\mathbf{W}}))^2} \mathbb{E}\left[\left(\int_0^t (U_{\mathbf{W}} - g(\mathbf{W}_s))^2 \mathrm{d}s\right)^\kappa \mid U_{\mathbf{W}}, L_{\mathbf{W}}, \kappa\right]\right] \\
&= \sum_{k=0}^{\infty} \frac{t^k}{p(k \mid U_{\mathbf{W}}, L_{\mathbf{W}}) k!^2} \mathbb{E}\left[\left(\int_0^t (U_{\mathbf{W}} - g(\mathbf{W}_s))^2 \mathrm{d}s\right)^k \mid U_{\mathbf{W}}, L_{\mathbf{W}}\right] .
\end{aligned}$$

Fubini's theorem and dominated convergence are used above (valid since the integrands are positive a.s.). (22) is obtained using the following result (which can be easily proved using Jensen's inequality). Let $f_i > 0$ for $i = 1, 2, \ldots$. Then the sequence of p_is which minimize $\sum_{i=0}^{\infty} f_i/p_i$ under the constraint $\sum p_i = 1$ is given by $p_i = \sqrt{f_i}/\sum \sqrt{f_i}$.

Appendix C: Proof of Theorem 2

GPE-1$\leq e^{-L_{\mathbf{W}} t}$ so that the result holds if $\mathbb{E}[e^{-L_{\mathbf{W}} t}] < \infty$, where the expectation is w.r.t. a d-dimensional Brownian bridge from $\mathbf{x}$ at time 0 to $\mathbf{y}$ at time t. However

$$\begin{aligned}
\mathbb{E}[e^{-L_{\mathbf{W}} t}] &= \int_0^\infty \mathrm{P}[e^{-L_{\mathbf{W}}} > w] \mathrm{d}w \\
&= \int_0^\infty \mathrm{P}[L_{\mathbf{W}} < -\log w] \mathrm{d}w \leq \int_0^\infty \mathrm{P}[\delta \sum_{i=1}^{d} (1 + M_i) > \log w] \mathrm{d}w
\end{aligned}$$

where $M_i = \sup_{0 \leq s \leq t} |W_i|$ using the the growth bound in Theorem 2. Furthermore,

$$\begin{aligned}
\int_0^\infty \mathrm{P}[\delta \sum_{i=1}^{d} (1 + M_i) > \log w] \mathrm{d}w &\leq \int_0^\infty \sum_{i=1}^{d} \mathrm{P}[\delta(1 + M_i) > d^{-1} \log w] \mathrm{d}w \\
&= \int_0^\infty \sum_{i=1}^{d} \mathrm{P}[M_i > (d\delta)^{-1} \log w - 1] \mathrm{d}w \,.
\end{aligned}$$

It remains therefore to bound the d integrals on the right hand side of this expression. However from the Bachelier-Levy formula for hitting times for Brownian motion and bridges,

$$\mathrm{P}[M_i > v] \leq \exp\left\{-2(v - \max\{x_i, y_i\})^2/t\right\} + \exp\left\{-2(\min\{x_i, y_i + v\})^2/t\right\}$$

and so

$$\begin{aligned}\mathrm{P}[M_i > (d\delta)^{-1}\log w - 1] &\leq \exp\{-2((d\delta)^{-1}\log w - 1) - \max\{x_i, y_i\})^2/t\} \\ &+ \exp\{-2(\min\{x_i, y_i + (d\delta)^{-1}\log w - 1)\})^2/t\}\end{aligned}$$

which recedes like $w^{-k\log w}$ as $w \to \infty$ thus concluding the proof.

Appendix D: EPPF and ESPF for Example 1

EPPF generates the new particles according to the following procedure:

(i) choose one of the current particles $x_i^{(k_{i,j})}$, where particle j is chosen w.p. $\beta_i^{(j)}$;

(ii) propose $x_{t_{i+1}}$ from Normal with mean $x_i^{(k_{i,j})}$, and variance Δ_i;

(iii) accept this proposal w.p. $\exp(-\cos(x_{t_{i+1}}) - 1)$; if proposal is rejected return to (i).

(iv) accept this proposal with probability

$$\mathbb{E}\left[\exp\left\{-\int_0^{\Delta_i}\phi(W_s)\mathrm{d}s\right\}\right],$$

where expectation is with respect to the law of a Brownian Bridge from $W_0 = x_i^{(k_{i,j})}$ and $W_{\Delta_i} = x_{t_{i+1}}$, and ϕ is given in (7). If the proposal is rejected return to (i), otherwise $x_{t_{i+1}}$ is the new particle at time t_{i+1} with weight $w_{i+1} = f(y_{i+1}|x_{i+1})$.

(iv) is performed using retrospective sampling as described in Beskos et al. (2006a).

ESPF proceeds as above but with steps (i) and (ii) replaced by the step

(i') propose $(x_i^{(k_{i,j})}, x_{t_{i+1}})$ according to the density proportional to

$$\exp\left\{-\cos(x_i^{(k_{i,j})}) - \frac{\left(y_{i+1} - x_i^{(k_{i,j})}\right)^2}{2(\sigma^2 + (\Delta_i))}\right\}\exp\left\{-\frac{(x_{t_{i+1}} - \eta)^2}{2\tau^2}\right\}$$

where $\eta = (\sigma^2 + \Delta_i)^{-1}(x_i^{(k_{i,j})}\sigma^2 + \Delta_i y_{i+1})$, and $\tau = \sigma^2\Delta_i/(\sigma^2 + \Delta_i)$.

The algorithm is repeated until N values for $x_{t_{i+1}}$ are accepted, each with weight $1/N$.

Appendix E: Proposal Distribution for Example 1

Consider a diffusion satisfying SDE (1), with $d = 1$ for simplicity. The Ozaki approximation of this SDE is based on a first order Taylor expansion of the drift about some value x. For the sine diffusion of Example 1, we get the following approximating SDE

$$\mathrm{d}\tilde{X}_s = -\cos(x)[x - \tan(x) - \tilde{X}_s]\,\mathrm{d}s + \mathrm{d}B_s.$$

So $\tilde{X}_s - (x - \tan(x))$ is an OU process as defined in Example 2 with $\rho = \cos(x)$ and $\sigma = 1$. To calculate $q(x_{t_{i+1}}|x_i^{(j)}, y_{i+1})$ we compute the product of the transition density given by the Ozaki approximation about $x = x_i^{(j)}$ and the likelihood function $f(y_{i+1}|x_{t_{i+1}})$. Defining $\tau^2 = (1-\exp\{-2\cos(x_i^{(j)})\Delta_i\})/(2\cos(x_i^{(j)}))$, and $\eta = x_i^{(j)} - \tan(x_i^{(j)})(1-\exp\{-\cos(x_i^{(j)})\Delta_i\})$ we get that $q(x_{t_{i+1}}|x_i^{(j)}, y_{i+1})$ is Normal with mean $(\eta\sigma^2 + y_{i+1}\tau^2)/(\tau^2\sigma^2)$ and variance $\eta^2\tau^2/(\eta^2 + \tau^2)$. Furthermore we calculate $\beta_i^{(j)} \propto w_i^{(j)}\mathcal{N}_{\tau^2+\sigma^2}(y_{i+1} - \eta)$.

Appendix F: Central Limit Theorem

For notational simplicity, we consider a special case of our particle filter, chosen to resemble those considered in Chopin (2004). We choose our proposal density for time t_{i+1} to have $\beta_j = w_i^{(j)}$; and we assume iid sampling of $X_{t_i}^{(j)}$ in step PF1. The particle filter of Chopin (2004) splits up simulating particles at time t_{i+1} into (i) a resampling of particles at time t_i; and (ii) a propagation of each of these particles to time t_{i+1}. Our assumption of iid sampling is equivalent to the multinomial resampling case of Chopin (2004). (The conditions for the central limit theorem are the same if the residual sampling methods of Liu and Chen (1998), but the variances differ.) For simplicity we consider observation model (A) or (B), though the result extends easily to observation model (C).

Let $\theta_i^{(j)} = (x_{t_i}^{(j)}, x_{t_{i-1}}^{(k_{i,j})})$, where $k_{i,j}$ is the index sampled in step PF1 when simulating the j particle at time t_i and $\theta_i^{(j)}$ is the jth particle at time t_i together with the particle at time t_{i-1} from which it is descended. Also let $\mathbb{E}_{\theta_i}$ denote conditional expectation given θ_i. Similarly, let $\mu_i(\theta_i) = \mu_g(x_{i-1}, x_i, t_{i-1}, t_i)$, and denote by R_i the unbiased estimator of $\mu_i(\theta_i)$, i.e. $\mathbb{E}[R_i] = \mu_i(\theta_i)$. An important quantity is $\sigma_i^2(\theta_i) = \mathrm{Var}(R_i)$.

We define $\mathbb{E}_i[\varphi]$ and $\mathrm{Var}_i(\varphi)$ to be the posterior mean and variance of an arbitrary function $\varphi(\theta)$ at time i, and consider Particle Filter estimates of $\mathbb{E}_i[\varphi]$. Let $\tilde{\pi}_i(\theta_i)$ be the density $p(x_{t_{i-1}}|y_{1:i-1})q(x_{t_i}|x_{t_{i-1}})$. Finally define $\mathbb{E}_{q_i}[\varphi]$ and $\mathrm{Var}_{q_i}(\varphi)$ to be shorthand for the conditional expectation and variance of $\varphi(\theta_i)$ with respect to $q(x_{t_i}|x_{t_{i-1}})$ (which are functions of $x_{t_{i-1}}$). We denote $\|\cdot\|$ to be the Euclidean norm and define recursively Φ_i to be the set of measurable functions φ such that for some $\delta > 0$ $\mathbb{E}_{\tilde{\pi}_i}[\|h_i R_i \varphi\|^{2+\delta}] < \infty$, and that the function $x_{t_{i-1}} \mapsto \mathbb{E}_{q_i}[h_i\mu_i\varphi]$ is in Φ_{i-1}.

Theorem 3. *Consider a function φ; define $\tilde{V}_0 = Var_{\tilde{\pi}(x_0)}(\varphi)$, and by induction:*

$$\tilde{V}_i(\varphi) = \hat{V}_{i-1}\left\{\mathbb{E}_{q_i}[\varphi]\right\} + \mathbb{E}_{i-1}\left\{Var_{q_i}(\varphi)\right\}, \text{ for } i > 0, \tag{27}$$

$$V_i(\varphi) = \frac{\tilde{V}_i\left\{\mu_i h_i \cdot (\varphi - \mathbb{E}_i[\varphi])\right\} + E_{\tilde{\pi}_i}((\varphi - \mathbb{E}_i[\varphi])^2 \sigma_i^2 h_i^2)}{E_{\tilde{\pi}_i}(\mu_i h_i)^2}, \text{ for } i \geq 0, \tag{28}$$

$$\hat{V}_i(\varphi) = V_i(\phi) + Var_i(\phi), \text{ for } i \geq 0. \tag{29}$$

Then if for all i (C1) $x_{t_i} \mapsto 1$ belongs to Φ_i; (C2) $\mathbb{E}_{\tilde{\pi}_i}[h_i^2 \sigma_i^2] < \infty$; and (C3) $\mathbb{E}_{\tilde{\pi}_i}[\sigma_i \varphi h_i]^{2+\delta} < \infty$ for some $\delta > 0$; then for any $\varphi \in \Phi_i$, $\mathbb{E}_i[\varphi]$ and $V_i(\varphi)$ are finite and we have the following convergence in distribution as the number of particles, N, tends to infinity:

$$N^{1/2}\left\{\frac{\sum_{j=1}^N w_i^{(j)} \varphi(x_{t_i}^{(j)})}{\sum_{j=1}^N w_i^{(j)}} - \mathbb{E}_i[\varphi]\right\} \rightarrow \mathcal{N}(0, V_i(\varphi))$$

Comment Equations (27)–(29) refer to the changes in variance of the weighted particles due to the propagation, weighting and resampling stages at iteration i. Only (28) differs from the respective result in Chopin (2004), and this is due to the second term on the right-hand side, which represents the increase in variance due to the randomness of the weights. Condition C1 is taken from Chopin (2004) and applies to standard particle filters; conditions C2 and C3 are new and are conditions bounding the variance of the random weights which ensures that $V_i(\varphi)$ is finite.

Proof. We adapt the induction proof in Chopin (2004), considering in turn the propagation, weighting and resampling steps Our filter differs from the standard particle filter only in terms of the weighting step; and therefore we need only to adapt the result of Lemma A2 in Chopin (2004). In fact, (27) and (29) are identical to the corresponding quantities in Chopin (2004), therefore it remains to show (28). We define the constant $K = \mathbb{E}_{\tilde{\pi}_i}[R_i h_i]$ and $\varphi^* = R_i h_i(\varphi - E_i(\varphi))/K$. Within the enlarged signal space framework, we can apply Equation (4) of Chopin (2004), to give:

$$V_i(\varphi) \;=\; \tilde{V}_i\left(\varphi^*\right) = \hat{V}_{i-1}\left\{\mathbb{E}_{q_i}[\varphi^*]\right\} + \mathbb{E}_{i-1}\left\{\mathrm{Var}_{q_i}(\varphi^*)\right\}.$$

Now we can calculate $\mathbb{E}_{q_i}[\varphi^*]$ by first taking expection over the auxiliary variables (conditional on θ_i). This gives $\mathbb{E}_{q_i}[\varphi^*] = \mathbb{E}_{q_i}[\mu_i h_i(\varphi - E_i(\varphi))/K]$. Similarly we get

$$\begin{aligned} \mathrm{Var}_{q_i}(\varphi^*) &= \mathrm{Var}_{q_i}(\mathbb{E}[R_i h_i(\varphi - E_i(\varphi))/K]) + \mathbb{E}_{q_i}(\mathrm{Var}[R_i h_i(\varphi - E_i(\varphi))/K]) \quad (30) \\ &= \mathrm{Var}_{q_i}(\mu_i h_i(\varphi - E_i(\varphi))/K) + \mathbb{E}_{q_i}(\sigma_i^2 h_i^2(\varphi - E_i(\varphi))^2/K^2). \quad (31) \end{aligned}$$

(Here the expectation and variance in (30) are w.r.t. the auxiliary variables). Combining these results gives (28). The regularity conditions (C1) - (C3) translate directly also. □

References

Aït-Sahalia, Y. (2004) Closed-form likelihood expansions for multivariate diffusions. Working paper, available from http://www.princeton.edu/∼yacine/research.htm.

Beskos, A., Papaspiliopoulos, O. and Roberts, G. O. (2005a) Monte Carlo maximum likelihood estimation for discretely observed diffusion processes. Submitted.

— (2005b) A new factorisation of diffusion measure and finite sample path constructions. Submitted.

— (2006a) Retrospective exact simulation of diffusion sample paths with applications. *Bernoulli*, **12**, 1077–1098.

Beskos, A., Papaspiliopoulos, O., Roberts, G. O. and Fearnhead, P. (2006b) Exact and efficient likelihood–based inference for discretely observed diffusions (with discussion). *Journal of the Royal Statistical Society, Series B*.

Beskos, A. and Roberts, G. O. (2005) Exact simulation of diffusions. *Annals of Applied Probability*, **15**, 2422–2444.

Carpenter, J., Clifford, P. and Fearnhead, P. (1999) An improved particle filter for nonlinear problems. *IEE proceedings-Radar, Sonar and Navigation*, **146**, 2–7.

Chib, S., Pitt, M. K. and Shephard, N. (2006) Likelihood based inference for diffusion driven state space models. *Submitted*.

Chopin, N. (2004) Central limit theorem for sequential Monte Carlo methods and its application to Bayesian inference. *The Annals of Statistics*, **32**, 2385–2411.

Crisan, D. (2001) Particle filters - a theoretical perspective. In *Sequential Monte Carlo Methods in Practice* (eds. A. Doucet, N. de Freitas and N. gordon), 17–41. Springer–Verlag; New York.

Crisan, D., Del Moral, P. and Lyons, T. J. (1999) Interacting particle systems approximations of the Kushner-Stratonovich equation. *Advances in Applied Probability*, **31**, 819–838.

Dacunha-Castelle, D. and Florens-Zmirou, D. (1986) Estimation of the coefficients of a diffusion from discrete observations. *Stochastics*, **19**, 263–284.

Dassios, A. and Jang, H.-W. (2005) Kalman-Bucy filtering for linear systems driven by the Cox process with shot noise intensity and its application to the pricing of reinsurance contracts. *J. Appl. Probab.*, **42**, 93–107.

Del Moral, P. and Guionnet, A. (2001) On the stability of interactin processes with applications to filtering and genetic algorithms. *Ann. Inst. of H. Poincaré Probab. Statist.*

Del Moral, P., Jacod, J. and Protter, P. (2001) The Monte-Carlo method for filtering with discrete-time observations. *Probab. Theory Related Fields*, **120**, 346–368.

Del Moral, P. and Miclo, L. (2000) *Branching and interacting particle systems. Approximations of Feymann-Kac formulae with applicationc to non-linear filtering.*

Doucet, A., de Freitas, N. and Gordon, N. (2001) An introduction to sequential Monte Carlo methods. In *Sequential Monte Carlo methods in practice*, Stat. Eng. Inf. Sci., 3–14. New York: Springer.

Duffie, D. and Singleton, K. J. (1999) Modeling term structures of defaultable bonds. *The Review of Financial Studies*, **12**, 687–720.

Durham, G. B. and Gallant, A. R. (2002) Numerical techniques for maximum likelihood estimation of continuous-time diffusion processes. *J. Bus. Econom. Statist.*, **20**, 297–338. With comments and a reply by the authors.

Engel, R. F. (2000) The econometrics of ultra-high-frequency data. *Econometrica*, **68**, 1–22.

Golightly, A. and Wilkinson, D. J. (2006) Bayesian sequential inference for nonlinear multivariate diffusions. *Statistics and Computing*, **16**, 323–338.

Gordon, N., Salmond, D. and Smith, A. F. M. (1993) Novel approach to nonlinear/non-Gaussian Bayesian state estimation. *IEE proceedings-F*, **140**, 107–113.

Ionides, E. (2003) Inference and filtering for partially observed diffusion processes via sequential Monte Carlo. Working paper available from http://www.stat.lsa.umich.edu/∼ionides/pubs/WorkingPaper-filters.pdf.

Kitagawa, G. (1996) Monte Carlo filter and smoother for non-Gaussian nonlinear state space models. *Journal of Computational and Graphical Statistics*, **5**, 1–25.

Kou, S. C., Xie, X. S. and Liu, J. S. (2005) Bayesian analysis of single-molecule experimental data. *J. Roy. Statist. Soc. Ser. C*, **54**, 469–506.

Künsch, H. R. (2005) Monte Carlo filters:Algorithms and theoretical analysis. *Annals of Statistics*, **33**, 1983–2021.

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

Nicolau, J. (2002) A new technique for simulating the likelihood of stochastic differential equations. *The Econometrics Journal*, **5**, 91–103.

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