Filtering systems of coupled stochastic differential equations partially observed at high frequency

Paul Fearnhead[*], Omiros Papaspiliopoulos[†], Gareth O. Roberts[†]
and Andrew Stuart [‡]

August 1, 2007

Abstract

We consider online analysis of systems of stochastic differential equations (SDEs), from high-frequency data. The class of SDEs we focus on have constant volatility and a drift function that is of gradient form. For these models we present a particle filter that is able to analyse the full data, but whose computational cost does not increase as the frequency of the data increases. The method is based on novel extensions of the exact algorithm for simulation and inference of diffusions, and the filters do not need to introduce any approximations through time-discretisation of the process. The new methods have important practical and theoretical advantages over existing filtering methods for this problem. We demonstrate our method on a number of simulated examples, including two motivated by molecular dynamics.

Keywords : Diffusions, Gaussian process, Kalman filter, change of measure, Poisson estimator, Auxiliary particle filter, negative weights

1   Motivation

We study large stochastic systems modelled via coupled Stochastic Differential Equations (SDEs). In particular, let Z be a d-dimensional process containing all states of the system. We model Z as a strong Markov process defined via the solution of an SDE of the form

$$\mathrm{d}Z_s = -\Sigma \nabla A(Z_s)\,\mathrm{d}s + \sqrt{2\Sigma}\mathrm{d}B_s\,, \quad s \in [0, T]\,, \quad Z_0 = z\,, \tag{1}$$

---

[*]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`

[‡]Department of Mathematics, Warwick University, U.K., email: `A.M.Stuart@warwick.ac.uk`

1

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. (Refer to Appendix A for a brief summary of the fairly standard linear algebra notation we will use in this paper). The first expression in the equation is known as the drift and the second as the diffusion coefficient. We assume standard conditions on A which ensure the existence and uniqueness of a solution of (1); these are summarised in Appendix B for completeness. We have chosen a parametrisation according to which Σ determines the speed of the process; if the function $\rho(u) \propto \exp\{-A(u)\}$ is integrable, then Z is a reversible Markov process with invariant density proportional to ρ, for any non-zero Σ. Generalizations of this model structure are discussed in Section 5.

Models such as that appearing in equation (1) are standard in the analysis of many physical systems. For example, in the context of molecular dynamics A is a potential energy, describing interactions amongst components of the system, and the noise term models thermal activation. In this context the model is often referred to as *Brownian dynamics* or as the *Smolochouski equation*; it arises as the high friction limit of the second order Langevin equation; see for example Gardiner (1985) for a book-length treatment of this modelling approach and several references.

Typically, the SDE specified in (1) cannot be solved analytically, and its transition density is intractable. Nevertheless, the class of SDEs in (1) is very appealing from a computational point of view since it permits the exploration of its finite dimensional distributions via exact and efficient Monte Carlo methods, which do not rely on approximate time-discretisations of the SDE. In particular, Beskos *et al.* (2005) introduced a rejection sampling algorithm for simulating Z_t conditionally on $Z_0 = z$. More recently Fearnhead *et al.* (2006) introduced a general importance sampling scheme for this conditional distribution, where values Z_t are proposed from some tractable distribution (e.g. a linearisation of the SDE) and are appropriately weighted. Beskos *et al.* (2006b) introduced a collection of techniques for likelihood inference for the parameters of (1) based on discrete time observations of pairs of (Y, X).

In this paper, we are interested in the case where Z is decomposed as $Z = (Y, X)$, $Y \in R^{d_1}, X \in R^{d_2}, d_1 + d_2 = d$, Y being the observable and X the unobservable part of the process. We correspondingly block Σ in terms of Σ_1, Σ_2 and Σ_{12}. In the context of molecular dynamics we might think of observing some subset of the configuration of the system (one atom in a crystal, or a side chain in a biomolecule for example) and attempting to make inference about the remainder of the system. We address the problem of on-line inference for the unobserved process X given observations of Y at times $0 = t_0 < t_1 < \ldots < t_n = T$. The posterior density of X at an observation time given all data observed up to that time is known as the filtering density. We are particularly interested in solving the filtering problem when Y is observed at high frequency.

One possible solution to the filtering problem for diffusions is provided by the methodology proposed in Fearnhead *et al.* (2006). That approach can be used to estimate the

2

filtering distributions at each observation time t_j, i.e. the distributions of X_{t_j} given all available data up to time t_j. Specifically, this filtering problem falls into the "Observation regime B" according to the terminology of that paper. In this paper we generalise substantially the particle filtering methodology of Fearnhead *et al.* (2006), and we provide methods which are well-suited for high frequency data. A byproduct of our work is an important extension of the simulation methodology of Beskos *et al.* (2006a, 2005) and Fearnhead *et al.* (2006), appropriate for simulating partially observed processes.

It is known (see for example Del Moral and Miclo, 2000a) that traditional methods for continuous time filtering typically suffer from two particular problems. Firstly, they often require a time-discretised approximation to the continuous-time model to be considered. Furthermore, since they are constructed using one step look-ahead filtering rules, their computational cost inevitably depends on the length of that "step". Since the step size is usually dictated by the time interval between consecutive data points, this leads to methods which are particularly sensitive to the data frequency, and which often become infeasible for high-frequency data.

Whilst Fearnhead *et al.* (2006) provides a filter which is unbiased with respect to discretisation error, the solution proposed is not ideally suited to high frequency data. The key contribution of this paper is to provide a truly continuous-time filter which is robust to the data frequency, and which is ideally suited to on-line use.

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. We design particle filters which at each time s_i estimate the filtering distribution of X_{s_i} and the fixed-lag smoothing distribution (see for example Doucet *et al.*, 2000, for this terminology) of the path, i.e. the distribution of $(X_s, s \in [s_{i-1}, s_i])$ conditionally on all available data up to time s_i, for each $i = 1, \ldots, m$; see Figure 1 for an illustration. Our methods involve simulation of (Y, X) at Poisson-distributed times on $[s_{i-1}, s_i]$ according to a Kalman smoother and associating each such skeleton with appropriate importance weights. A key feature of the method is that the number of points at which we simulate (Y, X) does not increase as the frequency of data increases.

This generalises Fearnhead *et al.* (2006) since we can take $m = n$ and $s_i = t_i$ for all i. However, the generalization we propose allows inference about X conditionally on batches of data. There are major computational advantages of this new approach. In particular, if you ignore the computational cost of accessing the data, or of calculating simple summaries such as the minimum and maximum of the data, then the cost of our proposed filter does not increase with the frequency of the observations. By comparison the cost of applying the filter of Fearnhead *et al.* (2006), or filters that use time-discretisation of the SDE (see for example Crisan and Lyons, 1999; Crisan *et al.*, 1999; Del Moral *et al.*, 2001) to the full data would increase linearly with the number of observations.

In practice, current filtering methods for such high-frequency methods would have to analyse only a subset of the data (obtained by sub-sampling for example), to avoid such an increasing computational cost. We emphasise that our filter obtains a computational
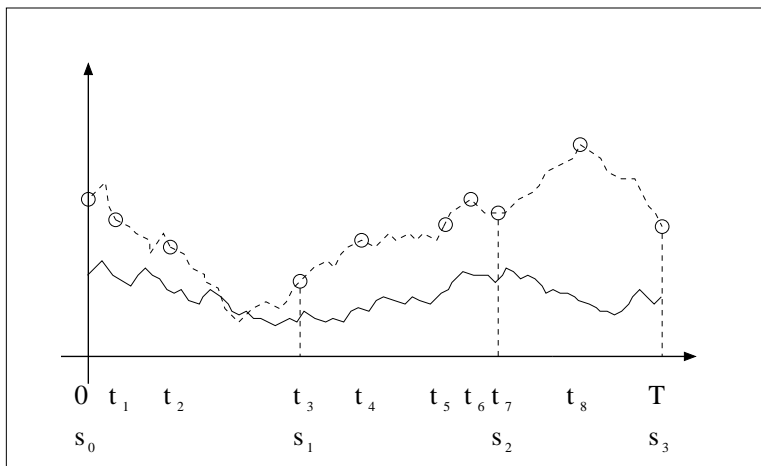
3

Figure 1: Observations from Y (denoted by circles) at times, $0, t_1, \ldots, t_8, T$; with dashed lines we denote the unobserved values of Y, and with solid line the X process, which is unobserved at all times. In this example there are ten observation four filtering times.

cost that is independent of the frequency of observations without any approximation or throwing away data. We simply exploit the structure of the model and recent advances in simulation of diffusion processes.

As an application of our methodology we consider inference for systems which evolve at different scales. A simple instance of such system arises when $\Sigma_2 = I_{d_2}$, $\Sigma_1 = (1/\epsilon)I_{d_1}$ and $\Sigma_{12} = 0$, where $\epsilon < 1$. In this case the Y variable evolves faster than X (see for example Figure 2 for a simulation from such model, which is considered later in the paper). Models which involve this scale separation are very intensively studied in a wide range of scientific fields where such dynamics have been observed, including atmospheric sciences, cell biology, molecular dynamics, material science and econometrics. The theory of averaging and homogenization is about the asymptotic (as $\epsilon \to 0$) analysis of such systems; indicative references of this field include the classic book by Bensoussan *et al.* (1978) and more recent expositions in Cioranescu and Donato (1999) and Pavliotis and Stuart (2008). Filtering for two-scale models has been considered for example in Papavasiliou and Kevrekidis (2007), Givon *et al.* (2006) and Vanden-Eijnden (2003), although from a perspective different from the one considered here.

The paper is organised as follows. In Section 2 we formulate the filtering problem, and present some key results that characterise the filtering distributions as a change of measure with respect to an appropriately defined Kalman smoother measure. We then present a range of particle filtering algorithms, including methods based on both rejection sampling and importance sampling. The latter builds on ideas in Fearnhead *et al.* (2006): we simulate weights that are assigned to each particle. These weights need to satisfy two

4

properties, that they are unbiased estimates of the true (but intractable) weight, and that they are positive. We present a new approach to ensuring positive weights, which use stopping times. This idea can be applied more widely than just to the filters considered here. In Section 4 we apply our filter to 3 simulated examples. These examples demonstrate the advantages that can be attained over existing filters. Furthermore, we see a robustness property of our filter to the choice of filtering times. The filter naturally adapts to the choice of filtering times by simulating more points between successive times as the filtering times become less frequent. This property compares with filters that discretise the SDE: where the choice of discretisation interval, which is equivalent to the interval between successive filtering times, can have a large impact on the accuracy of the filter and can be difficult to choose. Our paper concludes with a discussion about various extensions of our work.

2   Formulation of the filtering problem

Let $Z = (Y, X)$ be the stochastic process defined in (1), where Y is the observable and X the unobservable component of the process. In particular, we assume that $n + 1$ data points have been observed on $[0, T]$ at times $0 = t_0 < t_1 < \ldots < t_n = T$. We denote the set of available data between any two observation times $t_j < t_k$ by $y_{t_j:t_k} := (y_{t_j}, y_{t_{j+1}}, \ldots, y_{t_k})$.

Since X_0 is unobserved the initial condition in the SDE (1) will be partially unknown. Hence, to complete the model specification we will have to elicit a prior distribution for X_0. We will allow the prior to depend on y_0 and we will denote it by π_0. When (1) is ergodic, a natural choice is its invariant distribution: $\pi_{s_0}(x_0) \propto \exp\{-A(y_0, x_0)\}$. Our aim is to obtain a particle approximation of each filtering density of X at the collection of times $0 = s_0 < s_1 < \ldots < s_m = T$ which is a subset of the observation times. The ith filtering density, i.e. the density of X_{s_i} given $y_{0:s_i}$, evaluated at $X_{s_i} = x_i$, will be denoted by $\pi_{s_i}(x_i)$. It will be convenient to denote the joint density of $(X_{s_{i-1}}, X_{s_i})$ given $y_{0:s_i}$, evaluated at $X_{s_{i-1}} = x_{i-1}, X_{s_i} = x_i$, by $\pi_{s_i}(x_{i-1}, x_i)$, and the corresponding conditional density by $\pi_{s_i}(x_i \mid x_{i-1})$. When we write π_{s_i} without arguments then by default we refer to the filtering density $\pi_{s_i}(x_i)$. The term particle approximation is standard (see for example Doucet *et al.*, 2001) and it refers to a collection of N weighted "particles" which approximate the filtering distribution. Such samples are typically created by sequential application of importance or rejection sampling methods.

In general terms, if $\pi_{s_i}(x_i \mid x_{i-1})$ is available (up to normalising constant) then the standard particle filter machinery can be called upon to carry out the filtering estimation. However, there are two major challenges in the filtering problem we are considering in this paper. Firstly, it is typically impossible to derive in closed form these conditional densities due to the unavailability of the transition density for most diffusion processes. Secondly, we are conditioning on a collection of observed data between the successive states $X_{s_{i-1}}, X_{s_i}$, which is not standard. Additionally, the fact that we allow for general covariance Σ in (1)

5

adds some complications.

Hence, in order to address the filtering problem we will need first to find useful representations of $\pi_{s_i}(x_{i-1}, x_i)$. Our methodology is motivated from the fact that the law of (1) can be obtained as a change of measure from the law of the Gaussian process obtained by omitting the non-linear drift term from (1) (see also Del Moral and Miclo, 2000b, for a similarly motivated filtering approach). This will allow us to express the filtering distribution at s_i as a change of measure from the Gaussian process conditioned on $y_{0:s_i}$, i.e. from the Kalman smoother on $[0, s_i]$. We do this as described below.

Let $\mathbb{Q}(y_{s_{i-1}:s_i}, \pi_{s_{i-1}})$ denote the law of $Z = (Y, X)$ with dynamics according to (1), conditionally on $Y_{t_j} = y_{t_j}$ for all $s_{i-1} \le t_j \le s_i$, and with initial measure at time s_{i-1}, $X_{s_{i-1}} \sim \pi_{s_{i-1}}$. We emphasise that this is a probability measure on the space of paths of Z on $[s_{i-1}, s_i]$ which are consistent with the data. Then, it follows easily from the Markov property of (1) that the joint distribution of $(X_{s_{i-1}}, X_{s_i})$ conditionally upon $y_{0:s_i}$, is a marginal of $\mathbb{Q}(y_{s_{i-1}:s_i}, \pi_{s_{i-1}})$. On the same path space and for a probability measure $\nu_{s_{i-1}}$, we define a new stochastic process $W = (Y, X)$, with $Y \in R^{d_1}, X \in R^{d_2}$, as: $X_{s_{i-1}} \sim \nu_{s_{i-1}}$, $Y_{s_{i-1}} = y_{s_{i-1}}$, and conditionally on $X_{s_{i-1}}$, we set $W_s = W_{s_{i-1}} + \sqrt{2\Sigma} B_{s-s_{i-1}}$ for $s \ge s_{i-1}$, where B is a standard d-dimensional Brownian motion. Therefore, conditionally on $X_{s_{i-1}}$, W is a Gaussian process. Let $\mathbb{W}(y_{s_{i-1}:s_i}, \nu_{s_{i-1}})$ be the law of $W = (Y, X)$ conditionally upon $Y_{t_j} = y_{t_j}$ for all $s_{i-1} \le t_j < s_i$. Then, we have the following theorem, which forms the basis of our methodology. Its proof is given in Appendix C (see also Appendix A for notational conventions). In the theorem, and hereon throughout the paper, we assume that $\pi_{s_{i-1}}$ is absolutely continuous with respect to $\nu_{s_{i-1}}$.

Theorem 1. *Let $\pi_{s_{i-1}}, \nu_{s_{i-1}}$ be probability measures on R^{d_2}, with $\pi_{s_{i-1}}$ absolutely continuous with respect to $\nu_{s_{i-1}}$. Then, under mild technical and standard conditions (stated in Appendix C) $\mathbb{Q}(y_{s_{i-1}:s_i}, \pi_{s_{i-1}})$ is absolutely continuous with respect to $\mathbb{W}(y_{s_{i-1}:s_i}, \nu_{s_{i-1}})$. Additionally, under the following weak assumptions on the potential A,*

(A1) ∇A is continuously differentiable in all its arguments

(A2) there exists $l > -\infty$ such that

$$\phi(u) := \frac{1}{4}\big(\nabla A(u)^*\Sigma\nabla A(u) - 2\Sigma : \nabla\nabla A(u)\big) - l \ge 0\,, \quad \phi : R^d \to R \qquad (2)$$

the density between the two measures is proportional to

$$\frac{\mathrm{d}\pi_{s_{i-1}}}{\mathrm{d}\nu_{s_{i-1}}}(X_{s_{i-1}}) \times \exp\left\{-\frac{1}{2}(A(y_{s_i}, X_{s_i}) - A(y_{s_{i-1}}, X_{s_{i-1}})) - \int_{s_{i-1}}^{s_i} \phi(Y_s, X_s)\mathrm{d}s\right\}\,, \qquad (3)$$

where $Y_{t_j} = y_{t_j}$ for each observation time t_j on $[s_{i-1}, s_i]$. The constant of proportionality is a function of $y_{0:s_i}$.

6

Note that (A1) and (A2) are rather weak. For instance the first term in (A2) is a perfect square (since Σ is positive definite) and is quadratic in A, so this term very typically dominates the second term.

Therefore, a key representation of the filtering density $\pi_{s_i}(x_{i-1}, x_i)$ is as a marginal of the density specified in (3). Our plan is to construct Monte Carlo approximations of (3) which would then generate immediately approximations of the filtering densities.

Our first concern is that (3) is analytically intractable since the filtering distribution $\pi_{s_{i-1}}$, and its derivative with respect to $\nu_{s_{i-1}}$, will be analytically intractable (except trivially when $i = 1$). This is an issue in most non-linear/non-Gaussian filtering applications, not just the particular one we consider here (notable exceptions of tractable non-linear filters include Benes, 1981; Ferrante and Vidoni, 1998; Genon-Catalot, 2003). The standard solution to this problem is to replace $\pi_{s_{i-1}}$ by a particle approximation $\pi_{s_{i-1}}^{(N)}$, i.e. a set of N weighed particles, $\{x_{i-1}^{(j)}, w_{i-1}^{(j)}\}_{j=1}^N$, where $x_{i-1}^{(j)} \in R^{d_2}$, and $w_{i-1}^{(j)} \geq 0$, for all j. Effectively, this set forms an importance sampling approximation to $\pi_{s_{i-1}}$. Accordingly, we define $\nu_{s_{i-1}}^{(N)}$ to be a discrete probability measure with the same support as $\pi_{s_{i-1}}^{(N)}$, admitting a representation as $\{x_{i-1}^{(j)}, \beta_{i-1}^{(j)}\}_{j=1}^N$. Theorem 1 can be readily modified to give the density (up to proportionality) between $\mathbb{Q}(y_{s_{i-1}:s_i}, \pi_{s_{i-1}}^{(N)})$ and $\mathbb{W}(y_{s_{i-1}:s_i}, \nu_{s_{i-1}}^{(N)})$; let $(Y^{(j)}, X^{(j)})$ denote a path started from $X_{s_{i-1}} = x_{i-1}^{(j)}$, $j = 1, \ldots, N$, then (3) evaluated at $(Y^{(j)}, X^{(j)})$ becomes

$$\frac{w_{i-1}^{(j)}}{\beta_{i-1}^{(j)}} \times \exp\left\{ -\frac{1}{2}(A(y_{s_i}, X_{s_i}^{(j)}) - A(y_{s_{i-1}}, x_{i-1}^{(j)})) - \int_{s_{i-1}}^{s_i} \phi(Y_s^{(j)}, X_s^{(j)})\mathrm{d}s \right\}. \qquad (4)$$

This density yields $\pi_{s_i}^{(N)}(x_{i-1}, x_i)$ as a marginal, and it defines a transportation of the particle approximation from time s_{i-1} to time s_i. On the other hand, (4) is a density on the space of paths and its simulation involves in principle simulation of infinite dimensional objects. However, recent work (see in particular Beskos *et al.*, 2006a, 2005; Fearnhead *et al.*, 2006) has shown how to implement rejection and importance sampling for densities on the space of diffusion paths, using finite computational effort. The existing methods apply only to the simpler case where $\Sigma = I$, and crucially where there is no conditioning on intermediate observations. We now extend this simulation methodology to the partially observed case, thus providing rejection and importance sampling methods for (4). These Monte Carlo schemes are used to design two types of particle filters for approximating the sequence of the filtering densities. In Section 5 we comment on a further possibility for sampling from (4) which involves an interacting particle system.

3 Particle approximation of the filtering densities

We introduce two particle filtering schemes for solving the problem we described in Section 2. Throughout the section, we will assume that $\pi_{s_{i-1}}^{(N)}$ is an existing particle approximation

of $\pi_{s_{i-1}}$ obtained using importance or rejection sampling. At the core of the schemes lies a rejection and importance sampling methodology for partially observed diffusion processes, where a mixture of Kalman smoothers is used to generate proposals. We reiterate that a Monte Carlo approximation of the path density (4) immediately implies an approximation of any marginal of (4), in particular of $\pi_{s_i}^{(N)}(x_{i-1}, x_i)$ and $\pi_{s_i}^{(N)}(x_i)$. Although we will concentrate on those, we note that the methods we will present in this section also give a particle representation for the (fixed lag smoothing) distribution of the unobserved paths of (Y, X) between the observation times.

To simplify exposition we initially consider the case where ϕ in (2) is bounded. That is we assume that there exists some $r < \infty$, such that

$$0 < \phi(y, x) < r, \quad \text{for all } y \in R^{d_1}, x \in R^{d_2}. \tag{5}$$

The methods generalise easily when ϕ is only lower bounded, as is shown in Sections 3.3 and 3.4.

The following properties of the law $\mathbb{W}(y_{s_{i-1}:s_i}, \nu_{s_{i-1}})$ are requisite for our methodology. The proof, which we omit, can be derived using Kalman filter recursions, properties of Brownian motion and conditional properties of the multivariate Gaussian distribution.

Proposition 1. *Consider the process $W = (Y, X)$ defined in Section 2, and an arbitrary collection of times $t_0 < t_1 < \ldots, t_n$. Then, we have:*

(W1) *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 $Y_{t_i} \mid Y_{t_{i-1}} \sim N(Y_{t_{i-1}}, (t_i - t_{i-1})\Sigma_1)$.*

(W2) *For any $l = 1, \ldots, n$, the distribution of X_{t_l} conditionally on $X_{t_0} = x_0$ and $Y_{t_i} = y_{t_i}$, $0 \le i \le l$, is independent of Y_s for any $s > t_l$, and it has a Gaussian distribution,*

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

We will denote the density of this Gaussian distribution evaluated at a point $x_1 \in R^{d_2}$, by $\mathcal{G}_{t_l - t_0}(x_1 \mid x_0)$.

(W3) *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 $Y_{t_i} = y_{t_i}$, $0 \le i \le n$, is Gaussian, with mean and variance respectively*

$$\begin{aligned}
&\text{(mean)} \quad \frac{t_n - t_l}{t_n - t_0}(x_0 + \Sigma_{12}^* \Sigma_1^{-1}(y_{t_l} - y_{t_0})) + \frac{t_l - t_0}{t_n - t_0}(x_n + \Sigma_{12}^* \Sigma_1^{-1}(y_{t_l} - y_{t_n})) \\
&\text{(variance)} \quad 2\frac{(t_n - t_l)(t_l - t_0)}{t_n - t_0}(\Sigma_2 - \Sigma_{12}^* \Sigma_1^{-1} \Sigma_{12}).
\end{aligned} \tag{6}$$

(W4) *Assume that $Y_{t_i} = y_{t_i}$, $0 \le i \le n$, and that X has also been observed at two time points, $X_0 = x_0$, and $X_{t_k} = x_k$, for some $k \le n$. Consider a time $t_{l-1} < s < t_l$, for*

8

some $1 < l \le k$. 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,

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

and, conditionally on Y_s, X_s has the Gaussian law specified in (W3).

Moreover, we will require the following extension of $\mathbb{W}(y_{s_{i-1}:s_i}, \nu_{s_{i-1}})$. Let ν_{s_{i-1},s_i} be a probability measure on $R^{d_2} \times R^{d_2}$. Then we let $\mathbb{W}(y_{s_{i-1}:s_i}, \nu_{s_{i-1},s_i})$ denote the law of W conditionally on $Y_{t_j} = y_{t_j}$ for all $s_{i-1} \le t_j \le s_i$, where $(X_{s_{i-1}}, X_{s_i}) \sim \nu_{s_{i-1},s_i}$. When ν_{s_{i-1},s_i} is a point mass on a pair (x_{i-1}, x_i), we simply write $\mathbb{W}(y_{s_{i-1}:s_i}, x_{i-1}, x_i)$.

3.1 Fully adapted particle filter based on rejection sampling

We can produce a particle filter by sequential rejection sampling from (4). The output of each step of the algorithm is a set of equally weighted particles $\{x_{s_i}^{(j)}, 1/N\}_{j=1}^N$, for $i > 1$, which defines a particle approximation $\pi_{s_i}^{(N)}$ and it is used to define the next target density according to (4). In the particle filtering jargon, this is known as a fully adapted particle filter (Pitt and Shephard, 1999), since the particles are propagated according to the posterior distribution of the signal (i.e. the information in the data has been taken fully into account).

   We can perform such a rejection sampling by a careful extension of the Exact Algorithm of Beskos *et al.* (2006a, 2005). We design a rejection sampling algorithm for simulating from $\mathbb{Q}(y_{s_{i-1}:s_i}, \pi_{s_{i-1}}^{(N)})$ using proposals from a modification of the Kalman smoother measure on $[0, s_i]$. To this end, we require a further assumption on the potential function.

(A3) We assume that the function

$$f(x_{i-1}, x_i) := \mathcal{G}_{s_i - s_{i-1}}(x_i \mid x_{i-1}) \exp\left\{ -\frac{1}{2}(A(y_{s_i}, x_i) - A(y_{s_{i-1}}, x_{i-1})) \right\}, \quad (7)$$

is integrable in x_{i-1}, x_i, with respect to $\pi_{s_{i-1}}^{(N)} \bigotimes \mathbb{L}$, for all values of $y_{s_{i-1}}, y_{s_i}$, where $\mathbb{L}$ denotes the Lebesgue measure (on R^{d_2}), and $\mathcal{G}_{s_i - s_{i-1}}(x_i \mid x_{i-1})$ is defined in (W2) of Proposition 1.

This assumption is very weak and it will be met in all our examples.

   Let $\nu_{s_{i-1},s_i}^{(N)}$ be the probability measure on $R^{d_2} \times R^{d_2}$ with density with respect to $\pi_{s_{i-1}}^{(N)} \bigotimes \mathbb{L}$ proportional to (7). Then, it is easy to show (see Appendix D) that $\mathbb{Q}(y_{s_{i-1}:s_i}, \pi_{s_{i-1}}^{(N)})$ is absolutely continuous with respect to $\mathbb{W}(y_{s_{i-1}:s_i}, \nu_{s_{i-1},s_i}^{(N)})$ with density proportional to

$$\frac{\mathrm{d}\mathbb{Q}(y_{s_{i-1}:s_i}, \pi_{s_{i-1}}^{(N)})}{\mathrm{d}\mathbb{W}(y_{s_{i-1}:s_i}, \nu_{s_{i-1},s_i}^{(N)})}(Y, X) \propto \exp\left\{ -\int_{s_{i-1}}^{s_i} \phi(Y_s, X_s)\mathrm{d}s \right\} \le 1, \quad \text{since } \phi > 0, \quad (8)$$

9

with $Y_{t_j} = y_{t_j}$ for all $s_{i-1} \leq t_j \leq s_i$ in the above expression. Therefore, we can sample from $\mathbb{Q}(y_{s_{i-1}:s_i}, \pi_{s_{i-1}}^{(N)})$ by proposing paths (Y, X) from $\mathbb{W}(y_{s_{i-1}:s_i}, \nu_{s_{i-1},s_i}^{(N)})$ and accepting them with the probability given in (8). The retrospective simulation methodology introduced in Beskos *et al.* (2006a) provides an exact method for deciding on the acceptance of proposed paths (Y, X) with probability according to (8), by only simulating the process at s_{i-1} and s_i, and at a set of Poisson-distributed times on (s_{i-1}, s_i). Hence, using finite computational effort we can draw samples from $\pi_{s_i}^{(N)}(x_0, x_1)$, provided simulation from f in (7) is feasible (see the Remark below). However, the results of Beskos *et al.* (2006a) are restricted to the case where $\Sigma = I$, and crucially, when there are no intermediate observations of Y, therefore they have to be carefully extended. Below, we give the simulation algorithm which is appropriate for the situation we present here. We skip the proof of the validity of the algorithm, since this can be achieved by an argument similar to the one used in Section 2 of Beskos *et al.* (2006a), in conjunction with Theorem 1 and Proposition 1 of our paper.

Rejection sampling from $\pi_{s_i}^{(N)}(x_{i-1}, x_i)$

1. Simulate x_{i-1}, x_i according to the density in (7). Set $X_{s_{i-1}} = x_{i-1}, X_{s_i} = x_i$.

2. Let r be the upper bound of ϕ in (5). Simulate an integer $\kappa \sim \texttt{Poisson}(r(s_{i-1} - s_i))$.

3. Simulate κ points (ψ_j, υ_j) uniformly distributed on $(s_{i-1}, s_i) \times (0, 1)$, for $j = 1, \ldots, \kappa$, where the ψ_js are time-ordered. Conventionally, set $\psi_0 := s_{i-1}$.

4. Simulate iteratively pairs (Y_{ψ_j}, X_{ψ_j}) given $(Y_{\psi_{j-1}}, X_{\psi_{j-1}})$ and X_{s_i} according to the Gaussian distribution specified in (W4).

5. If $\upsilon_j > \phi(Y_{\psi_j}, X_{\psi_j})$ for all $j = 1, \ldots, \kappa$, then $(X_{s_{i-1}}, X_{s_i})$ is an exact draw from $\pi_{s_i}^{(N)}(x_{i-1}, x_i)$. Otherwise, return to 1.

Remark: Direct simulation from (7) might not be possible, but a rejection sampling scheme will typically be feasible, though great care is required to ensure that the computational costs of this scheme are not too large. Fortunately, the Lipschitz conditions on A can be utilised to construct an efficient adaptive rejection sampling, see Peluchetti (2007).

The advantage of the rejection sampling methodology is that it yields independent samples according to the filtering distributions. However, it has two potential drawbacks. The first relates with the computational effort required by the algorithm, since a certain proportion of the simulated samples will be rejected. This is magnified when further rejection sampling is used for simulating from (7). Secondly it avoids the use of Monte Carlo variance reduction techniques: such as the use of stratified (or no) resampling of

10

particles at time $s-1$. The rejection sampling algorithm always performs multinomial resampling of these particles, albeit from their correct conditional distribution given the new data.

3.2 Random weight particle filter by importance sampling

An alternative particle filter is obtained by sequential importance sampling from (4). In this scheme we have flexibility in the choice of proposal distribution. For the moment we concentrate on proposing from mixtures of Kalman smoothers, i.e. we propose paths from $\mathbb{W}(y_{s_{i-1}:s_i}, x_{i-1}, x_i)$, where the end-points (x_{i-1}, x_i) are distributed according to $\nu_{s_{i-1},s_i}^{(N)}$. We specify $\nu_{s_{i-1},s_i}^{(N)}$ in terms of its density with respect to $\pi_{s_i}^{(N)} \bigotimes \mathbb{L}$, which we will denote by $\nu_{s_{i-1},s_i}^{(N)}(x_{i-1}, x_i)$. This density evaluated at a pair $(x_{i-1}^{(j)}, x_i)$ takes the generic form

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

This proposal is equivalent to that used in the ASIR filter of Pitt and Shephard (1999). Note that simulating from this is achieved by simulating a particle $x_{i-1}^{(j)}$ with probability $\beta_{i-1}^{(j)}$, and then simulate x_i from $q_{s_i}(x_i \mid x_{i-1}^{(j)}, y_{0:s_i})$.

In practice we wish to choose this density so that it is easy to simulate from and it provides an approximation to $\pi_{s_i}^{(N)}(x_{i-1}, x_i)$. A simple approach is to approximate the SDE via the Euler discretisation. The Euler discretisation defines an approximate transition density over time interval $s_i - s_{i-1}$, which can be factorised as $p(y_{s_i}|x_{i-1}, y_{s_{i-1}})p(x_i|y_{s_i}, x_{i-1}, y_{s_{i-1}})$. We then define $\beta_{i-1}^{(j)} = p(y_{s_i}|x_{i-1}^{(j)}, y_{s_{i-1}})$ and $q_{s_i}(x_i \mid x_{i-1}^{(j)}, y_{0:s_i}) = p(x_i|y_{s_i}, x_{i-1}^{(j)}, y_{s_{i-1}})$. For alternative approaches for designing this proposal distribution for general state-space models see Pitt and Shephard (1999).

In order to carry out importance sampling we need to derive the likelihood ratio between the target and proposal measures. Let $(x_{i-1}^{(j)}, x_i)$ be a proposal from $\nu_{s_{i-1},s_i}^{(N)}$, and (Y, X) a path proposed from $\mathbb{W}(y_{s_{i-1}:s_i}, x_{i-1}^{(j)}, x_i)$. Note that it would be more consistent with the notation to write $(Y^{(j)}, X^{(j)})$, however this will be avoided to keep the formulae manageable. Appealing to a similar argument as the one we used in Section 3.1 (see Appendix D), we can easily show that the likelihood ratio evaluated at the proposed path is proportional to

$$\frac{w_{i-1}^{(j)}}{\beta_{i-1}^{(j)}} \frac{\mathcal{G}_{s_i-s_{i-1}}(x_i \mid x_{i-1}^{(j)})}{q_{s_i}(x_i \mid x_{i-1}^{(j)}, y_{0:s_i})} \exp\left\{-\frac{1}{2}(A(y_{s_i}, x_i) - A(y_{s_{i-1}}, x_{i-1}^{(j)}))\right\} \exp\left\{-\int_{s_{i-1}}^{s_i} \phi(Y_s, X_s)\mathrm{d}s\right\} , \tag{9}$$

where the constant of proportionality is a function of $y_{0:s_i}$. Therefore, each pair $(x_{i-1}^{(j)}, x_i)$ proposed from $\nu_{s_{i-1},s_i}^{(N)}$ will have to be weighted according to (9) in order to provide a sample from $\pi_{s_i}^{(N)}(x_{i-1}, x_i)$. However, the weight cannot be explicitly computed due to the last term in the product,

$$\exp\left\{-\int_{s_{i-1}}^{s_i} \phi(Y_s, X_s)\mathrm{d}s\right\} \tag{10}$$

11

which involves an integral over the unobserved path.

Fearnhead *et al.* (2006) introduced a generic methodology for filtering in the presence of intractable densities, according to which the intractable importance weights are replaced by positive unbiased estimators. This methodology is termed random weight importance sampling. They establish that this approach is equivalent to ordinary importance sampling on an expanded space which includes appropriately defined auxiliary variables.

We will adopt this paradigm and we will replace (10) by an unbiased estimator. A generic methodology exists for generating unbiased estimators of exponential functionals of stochastic processes, in particular of the type that appear in (9). This methodology is summarised in Appendix E; recent advances in this methodology were presented in Beskos *et al.* (2006b); Fearnhead *et al.* (2006); Jourdain and Sbai (2007). Notice that we are interested in simulating the weight under the assumption that (Y, X) are generated from $\mathbb{W}(y_{s_{i-1}:s_i}, x_{i-1}^{(j)}, x_i)$. Then, the construction of Appendix E in conjunction with Proposition 1, yields the following unbiased estimator (10):

$$e^{(\lambda-r)(s_i-s_{i-1})}\lambda^{-\kappa}\prod_{j=1}^{\kappa}(r-\phi(Y_{\psi_j}, X_{\psi_j})),\tag{11}$$

where $\kappa \sim \mathrm{Poisson}(\lambda(s_i - s_{i-1})$, $\lambda > 0$ is a user-specified constant which controls the variance of the estimator, the ψ_js are an ordered uniform sample on (s_{i-1}, s_i), and the pairs (Y_{ψ_j}, X_{ψ_j}) are simulated sequentially conditionally on $X_{s_i} = x_i$ and $X_{s_0} = x_{i-1}^{(j)}$, according to (W4). See Fearnhead *et al.* (2006) for discussion on the choice of λ, which typically will depend on the proposed pair $(x_{i-1}^{(j)}, x_i)$. Replacing (10) with (11) in (9) gives an unbiased estimate of the weight assigned to each proposed pair $(x_{i-1}^{(j)}, x_i)$; we will denote this weight by $\hat{w}_i^{(j)}$. Hence, we have the following generic particle filter.

Random weight particle filter

PF0: Simulate $x_0^{(j)}$, $j = 1, \ldots, N$, from ν_{s_0}, weight each value by $w_0^{(j)} = (\mathrm{d}\pi_{s_0}/\mathrm{d}\nu_{s_0})(x_0^{(j)})$. For $i = 1, \ldots, m$, for $j = 1, \ldots, N$:

PF1 Calculate the effective sample size of the $\{\beta_{i-1}^{(k)}\}_{k=1}^N$, $ESS := (\sum_{k=1}^N (\beta_{i-1}^{(k)})^2)^{-1}$. If $ESS < C$, for some fixed constant C, simulate $k_{i-1,j}$ from $p(k) \propto \beta_{i-1}^{(k)}$, $k = 1, \ldots, N$ and set $\delta_i^{(j)} = 1$; otherwise set $k_{i-1,j} = j$ and $\delta_i^{(j)} = \beta_{i-1}^{(j)}$.

PF2 Simulate $x_i^{(j)}$ from $q_{s_i}(x_i | x_{i-1}^{(k_{i,j})}, y_{0:s_i})$.

PF3 Generate an unbiased estimator $\hat{w}_i^{(j)}$ of (9).

PF4 Assign particle $x_i^{(j)}$ a weight $w_i^{(j)}$ equal to $\delta_i^{(j)} \hat{w}_i^{(j)}$.

12

Notice that our specification of the particle filter allows for the choice of whether to resample at each time step. This choice is based on the ESS of the particle weights (see e.g. Liu and Chen, 1998). Also note that to implement the random weight particle filter, calculation of the lower bound l in the definition of ϕ is (2) is not needed.

3.3 Generalizing the methods using layered Gaussian processes

We now assume that the function ϕ is unbounded from above. First see how this affects each of the methods we have proposed. The rejection sampling is still valid, in the sense that the likelihood ratio between the proposal and target measure is given by (8) and it is bounded by 1; this only requires that $\phi > 0$. However, the possibility to decide on the acceptance of a proposed path given only a finite skeleton of it, is no longer possible: the construction involves generating a Poisson process of rate equal to the upper bound of ϕ. On the other hand, the unbiased estimator of the importance sampling weights of Section 3.2 is valid regardless of the bound on ϕ. However, we can no longer guarantee that the weights will be positive; this is clear from (11). This is a serious problem, since we treat these weights as probabilities and use them to define the particle approximations of the filtering distributions.

A recent simulation construction, the layered Brownian bridge of Beskos *et al.* (2005), can be utilised to generalise the rejection and importance sampling methods of Sections 3.1 and 3.2, at extra computational cost. Let B be a Brownian motion conditioned on $B_0 = b_0, B_t = b_t$; this is known as the Brownian bridge. The construction of Beskos *et al.* (2005) allows the simulation of finite stochastic bounds, $\underline{b} < B_s \leq \bar{b}$ for all $s \in [0, t]$. These bounds depend on the time increment and b_0, b_t. Moreover, the construction allows the simulation of B at any collection of intermediate times, according to the Brownian bridge dynamics conditionally on the stochastic bounds.

We can use this approach to generate bounds for Y and X separately, on each time interval $[s_{i-1}, s_i]$. In particular, let (x_{i-1}, x_i) be an arbitrary pair of proposed values in either the importance or rejection sampling. The bounds for Y are constructed as follows. For each pair of consecutive observation times t_{j-1}, t_j, we write $Y_s, s \in [t_{j-1}, t_j]$ as $Y_s = \sqrt{2\Sigma_1} B_{s-t_{j-1}}$, where B is a Brownian bridge on $[0, t_j - t_{j-1}]$ starting from $\sqrt{2\Sigma_1}^{-1} y_{t_{j-1}}$ and finishing at $\sqrt{2\Sigma_1}^{-1} y_{t_j}$. Therefore, bounds on B imply bounds on Y. Notice that this has to be repeated for each pair of observation times on $[s_{i-1}, s_i]$. For bounding X we write for any $s \in [s_{i-1}, s_i]$,

$$X_s = \frac{s_{i-1} - s}{s_{i-1} - s_i} \Sigma_{12}^* \Sigma_1^{-1} (Y_s - y_{s_{i-1}}) + \frac{s - s_{i-1}}{s_i - s_{i-1}} \Sigma_{12}^* \Sigma_1^{-1} (Y_s - y_{s_i}) + \sqrt{2(\Sigma_2 - \Sigma_{12}^* \Sigma_1^{-1} \Sigma_{12})} B_{s-s_{i-1}}$$

where B is a Brownian bridge on $[0, s_{i-1} - s_i]$, starting from $\sqrt{2(\Sigma_2 - \Sigma_{12}^* \Sigma_1^{-1} \Sigma_{12})}^{-1} x_{i-1}$, and finishing at $\sqrt{2(\Sigma_2 - \Sigma_{12}^* \Sigma_1^{-1} \Sigma_{12})}^{-1} x_i$. Therefore, bounds on B and Y imply bounds on

13

X. Moreover, using the above representations we can simulate Y and X on $[s_{i-1}, s_i]$, conditionally on the bounds, by simulating the corresponding Brownian bridges and applying the appropriate transformation.

Hence, it is now possible to find a bound, $r(x_{i-1}, x_i)$ say, on $\phi(Y_s, X_s)$ for $s \in [s_{i-1}, s_i]$, and reduce the problem to the bounded case. In practice, bounding ϕ given bounds on (Y, X) can be nontrivial: see Appendix F.

3.4 Particle filtering with negative weights

We now consider an alternative implementation of the filter for unbounded ϕ which avoids using the layered Gaussian process. It has the advantage of simplicity and often speed, but at the expense of a slight reduction in the rate of convergence of the particle filter.

The basic idea is to construct probabilistic bounds on ϕ, conditional on the start and end of the state process and the observations. For calculating the weight of one particle at s_i, which corresponds to a state path from x_{i-1} at time s_{i-1} to x_i at time s_i, denote the bound by $r(x_{i-1}, x_i)$. The idea is to choose this such that

$$\mathrm{P}(\max(\phi(Y_t, X_t)) > r(x_{i-1}, x_i)) < \epsilon_N,$$

where the probability is calculated under $\mathbb{W}(y_{s_{i-1}:s_i}, x_{i-1}, x_i)$. We let the probability bound depend on the number of particles for reasons described below; thus $r(x_{i-1}, x_i)$ will also depend on N, but we suppress this in our notation. Bounds of this form can be constructed from probabilistic bounds on the (Y, X) path. As the law of these paths is given by a simple Gaussian process, good bounds on the (Y, X) path are simple to obtain. Furthermore, if $\phi(y, x)$ is a polynomial of order d in y and x, the bound $r(x_{i-1}, x_i)$ will only increase according to some power of $-\log(\epsilon_N)^{d/2}$.

Using this bound we then simulate a weight for the particle. Denote by $\mathcal{W}$ the weight assigned to the particle. We will have that $\mathrm{P}(\mathcal{W} < 0) < \epsilon_N$, as negative weights can only occur for paths for which the maximum value of ϕ exceed $r(x_{i-1}, x_i)$. Note that be choosing ϵ_N sufficiently small we can control the probability with which we observe negative weights.

If we implement the RWPF using such probabilistic bounds, we still need to consider how to deal with negative weights when they occur. We suggest two possibilities. The first is to replace negative weights with zero weights, or equivalently remove particles with negative weights. This approach will introduce a bias into the method, but by choosing ϵ_N sufficiently small we can control the bias. An informal argument suggests that is we choose $\epsilon_N = (N^{-\alpha})$ then the bias will be $O(N^{-\alpha})$. So for $\alpha > 1/2$ this bias will be asymptotially negligible as compared to the Monte Carlo error of the filter. Assuming $\phi(y, x)$ is polynomial of order d in y and x then such a choice will lead to a CPU cost of the filter that is $O(N \log(N)^{d/2})$, and thus the convergence rate in terms of CPU cost, C, will be $O(\log(C)^{d/2} C^{-1/2})$. This is an order $\log(C)^{d/2}$ worse than in the bounded ϕ case, but still substantially better than the $O(C^{-1/3})$ or $O(C^{-1/4})$ rates of convergence obtained using particle filters with discretisation (see the discussion in Fearnhead *et al.*, 2006).

14

The second approach is use extra simulation to remove negative weights. This can be done whilst ensuring the weights remain unbiased (up to a common constant of proportionality) using the following result.

Theorem 2. *Consider an infinite array of independent random variables $\mathcal{W}_k^{(j)}$ for $j = 1, \ldots, N$ and $k = 1, 2, \ldots$. 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 $K = \min\{l : S_l^{(j)} \geq 0 \text{ for all } j = 1, \ldots, N\}$. If $E(K) < \infty$, then

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

Proof: This is essentially Wald's identity (see for example Proposition 2.18 of Siegmund (1985)).

Thus we can implement an algorithm as follows.

Stopping-rule Simulation of Weights

1. Simulate a set of particles $(x_{i-1}^{(j)}, x_i^{(j)})$, for $j = 1, \ldots, N$, and calculate the bounds $r(x_{i-1}^{(j)}, x_i^{(j)})$.

2. Using these bounds, simulate weights $w_i^{(j)}$ for $j = 1, \ldots, N$.

3. While $\min_j\{w_i^{(j)}\} < 0$, simulate new weights $w_i^{*(j)}$ for each particle, and let $w_i^{(j)} = w_i^{(j)} + w_i^{*(j)}$.

Now providing the weights simulated in step (2) have finite sixth moments (which will hold for the generalised Poisson estimator of Fearnhead *et al.* (2006) using similar arguments to those used in that paper) then the weights simulated by this algorithm will have expectation proportional to the true weights, as required. The fact that the existence of sixth moments is sufficient follows by demonstrating that $K^* = \inf\{l; S_k^{(j)} \geq 0 \text{ for all } k \geq l, \text{ for all } j\}$ has finite expectation (since clearly $K \leq K^*$). However the finiteness of $E(K^*)$ follows from a standard argument using Markov inequality and a Borel-Cantelli Lemma.

The advantage of this approach over truncating is that it can be applied even when it is difficult to calculate $P(\max(\phi(Y_t, X_t)) > r(x_{i-1}, x_i))$ accurately, for your preferred choice of $r(x_{i-1}, x_i)$. In this case, the effect of choosing a poor bound is purely to increase the computational cost of the filter, rather than introducing (potentially) large biases into the algorithm.

15

4   Illustration of the methodology

We illustrate our methods on the following 3 examples. The first one is a linear SDE; for which the exact filtering densities can be calculated using the Kalman Filter Kalman and Bucy (1961). We use this example to evaluate the performance of our method and compare it to filters that require discretisation. We then consider two model problems taken from molecular dynamics based on a double-well potential (see e.g. Metzner *et al.*, 2006) and the Fixman potential (originally presented in Fixman, 1978).

In all cases we implemented the random weight particle filter (henceforth RWPF) using the stopping time idea of Theorem 2 to correct for negative weights. (Similar results are obtained using the layered Gaussian process but results are omitted for brevity.) Details of how approximate bounds on $\phi(x, y)$ were calculated is given in Appendix F. In all cases we used 1,000 particles and used an Euler approximation for $p(x_i, y_{s_i}|x_{i-1}, y_{s_{i-1}})$ to calculate the β_{i-1}s and the $q_{s_i}(x_i|x_{i-1}, y_{0:s_i})$ used in the filter. In simulating the random weight, we chose the mean number of points to be simulated using the suggestions in Fearnhead *et al.* (2006). Resampling for the particle filters was via the stratified sampling approach of Carpenter *et al.* (1999), and resampling was used when the ESS of the particle weights dropped below $N/2$. For simplicity we have considered two-dimensional systems, with $\Sigma_1 = 1/(2\epsilon)$, $\Sigma_2 = 1/2$ and $\Sigma_{12} = 0$. 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 different values of ϵ.

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 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 realisation of the process is given in Figure 2 with $\epsilon = 1/100$. We then applied the RWPF to analyse this data, using 10^5 observations. We chose 100 equally spaced filtering times. The filtered mean and regions of plus/minus two standard deviations are shown in Figure 2 together with the exact quantities as calculated by the Kalman filter. By eye, there is no noticeable difference between the Kalman filter and RWPF results.

We also compared the performance of the RWPF with a filter based on discretising 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 is approximated through an Euler approximation to the SDE. A standard particle filter can then be applied to the resulting discrete-time problem; in practice we implemented a fully adapted version of the ASIR filter Pitt and Shephard (1999). We implemented such a particle filter with 1000 particles, which we call a discrete-time particle filter (DPF). For this problem, after discretising time we have a simple linear-Gaussian state-space model, for which we can calculate the exact filtering distributions using the Kalman filter. We also looked at this
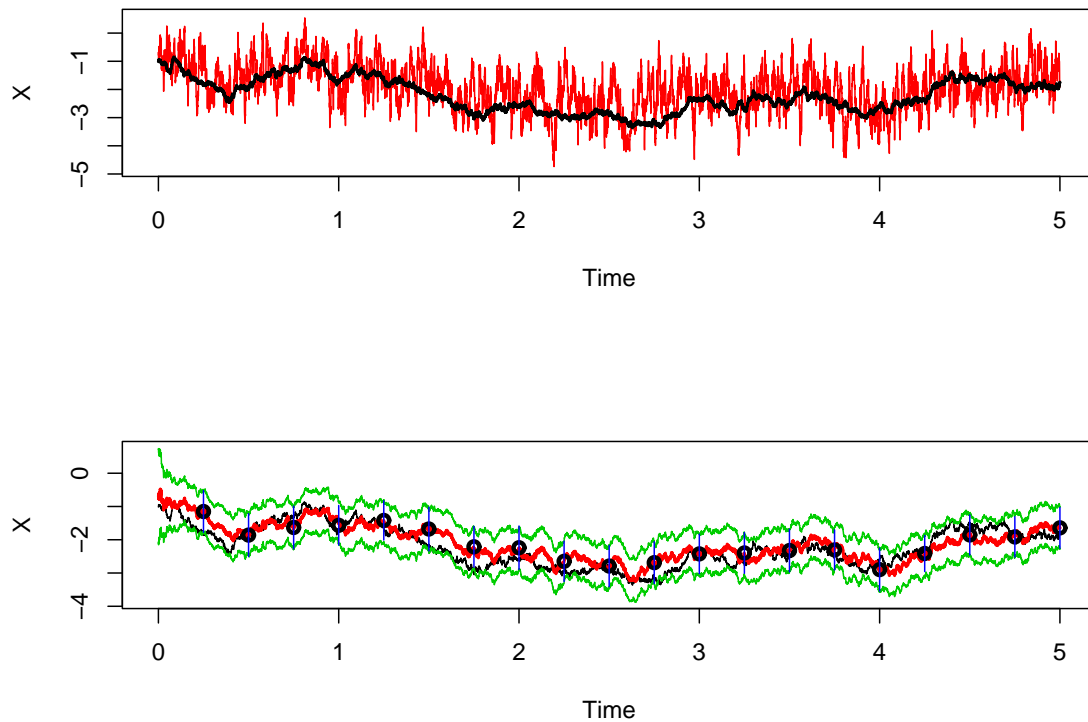
16

Figure 2: (Top) Simulated realisation of the OU process. The black line shows the (slow) unobserved state, and the red line shows the (fast) observed state. (Bottom) Unobserved state (black); true posterior means (red); RWPF estimates at 20 time points (blue circles). Uncertainty in the estimates are shown by regions of plus/minus two standard deviations: exact (green) and based on the RWPF (blue). The RWPF results were based on 100 filtering times.

17

approach, which we denote the discrete-time Kalman filter (DKF). The DKF is equivalent to the performance of DPF with an infinite number of particles.

A comparison of the three methods is shown in Figure 3, for different numbers of filtering times and different 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 RWPF gives a substantial reduction in MSE over the other two filters. Furthermore we see that the Monte Carlo error within the particle filter is small, as both DPF and DKF give almost identical results. The RWPF's performance is also robust to the number of filtering times, as it uses all the information in the data regardless, unlike DPF or DKF. Its performance is slightly worse for smaller number of filtering times, due to the increased Monte Carlo variation in simulating the weights in these cases. The computational cost of the RWPF is reasonably robust to the choice of the number of filtering times. For example, for $\epsilon = 1/100$ the total number of simulations per particle (equal to the number of filtering steps plus the number of points simulated in calculating the weights) ranges from 800 (300 filtering times) to 1,250 (1,000 filtering times) over the different choices; though would start to increase linearly as the number of filtering times increases beyond 1,000.

Note that a direct comparison of RWPF and DPF for the same number of filtering times is unfair – as the amount of simulation per particle for the DPF is just equal to the number of filtering times. However, even taking this into account (for $\epsilon = 1/100$ compare RWPF with 300 times versus PDF with 800 filtering times) the RWPF is substantially more accurate.

Example 2: Double well potential. A more challenging example, is specified by the following potential function, where we take $d_1 = d_2 = 1$

$$A(x, y) = q_1(y^2 - \mu)^2 + q_2(y - q_3 x)^2, \quad q_1, q_2 > 0, \mu, q_3 \in \mathbf{R}, \tag{12}$$

For our simulations we took $q_1 = 1$, $q_2 = 8$, $q_3 = 1$, $\mu = 1$ and $\epsilon = 1/100$. The potential produces a stationary distribution with two modes, at $(x, y) = (1, 1)$ and $(-1, -1)$.

Here we focus solely on the performance of the RWPF. We simulated data over two units of time, with 2×10^5 observations. Our simulated data was chosen to include a transition of the process between the two modes. We analysed the data using 500 filtering times. The results are shown in Figure 4. In this example we simulated the process at an average of 8 Poisson time points between each time-point; which suggests having more frequent filtering times would be preferable. However even in this case, resampling was only required at every other time-step.

Example 3: Fixman potential. The context here is the effect of constraining molecular dynamics by fixing bond lengths between atoms. Our model in this section is a mathematical idealization of this original set-up. The Fixman potential arises in the limit as $\epsilon \to 0$. In this case, the method of averaging shows that X converges weakly to the solution of the
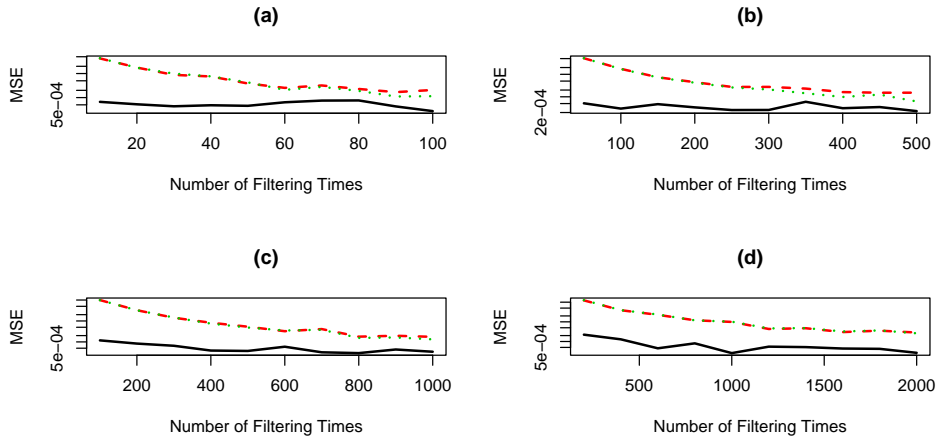
18

Figure 3: Comparison of RWPF (black), DPF (red dashed) and DKF (green dotted) at approximating the filtered mean. Results are for the mean square error relative to the truth, and for (a) $\epsilon = 1/10$; (b) $\epsilon = 1/25$; (c) $\epsilon = 1/100$; and (d) $\epsilon = 1/400$.
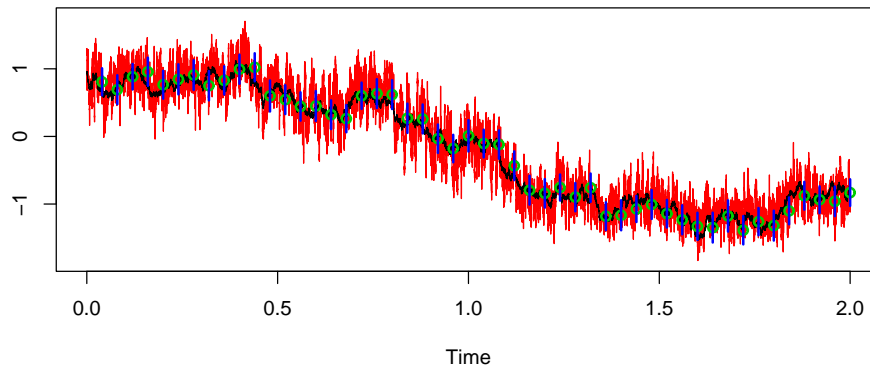

Figure 4: Results for the double well model. True state (black), observed process (red), filtered means (green circles) and regions plus/minus two standard errors (blue). The RWPF was run for 500 filtering times; but for clarity results at only 50 are shown.

19

SDE
$$dX_s = -\nabla\Psi(X_s)ds + \sqrt{2}dB_s, \quad s \in [0,T],$$

where B is a standard d_2-dimensional Brownian motion (independent of the one in (1)) and Ψ is the so-called *Fixman potential*. This is defined by

$$\exp\Big(-\Psi(x)\Big) = \int_{\mathcal{Y}} \exp\Big(-A(y,x)\Big)dy.$$

The primary mechanism inherent in the averaging procedure is that the projection of the drift in the $x-$coordinate is averaged against the invariant measure of the Y process, with X viewed as a fixed parameter. As such, the increments of Y are irrelevant, it is simply the "local" empirical measure for Y (found on time-scales long compared to ϵ and short compared to 1) which matters. In such a situation it is natural to study how well the proposed filter performs in the reconstruction of X given high frequency data in Y. Our filter can give a probabilistic reconstruction of the unobserved process for arbitrary values of ϵ and it is well-designed to handle high-frequency observations from Y.

A interesting family of problems when $d = 2$ motivated from molecular dynamics comes from taking

$$A(y,x) = \frac{1}{2}\big(y - m(x)\big)^2 q(x) + p(x)$$

where $q(x)$ is strictly positive. Then

$$\Psi(x) = p(x) + \frac{1}{2}\ln q(x).$$

We chose $m(x) = x$, $q(x) = 1/2$ and $p(x) = \sin(2x) + 2\sin(3x)$. In this setting x would represent an angle between atoms. The Fixman potential produces a model for x which has three local modes in $[0, 2\pi]$, which in order of size are at 5.7, 1.7 and 3.6 respectively. We simulated data with $\epsilon = 1/400$ for 2.5 time units, and we chose a simulation which include a transition between two modes.

In total 10^5 observations were simulated, and we analysed the data using 500 filtering times. Both the simulated data and the results of the RWPF are shown in Figure 5. Note that at some time points (such as the third time point where the filtered mean is shown in Figure 5), the filtering density is bimodal – each mode corresponding to the state being in a different mode of the stationary distribution of the model. For these settings the RWPF on average simulated the path at 0.7 points between successive filtering times. Resampling was used on average at every 10th filtering time point.

For comparison we ran the DPF with $1,000$ particles and $1,000$ filtering times. To compare the results of the two filters we obtained an accurate estimate of the filtering means using a DPF with $10,000$ particles and $5,000$ filtering steps. We then calculate the MSE of the two filter estimates with this "truth". The MSE was 0.0009 for the RWPF and 0.017 for the DPF, again showing a substantial increase in accuracy of the RWPF over a particle filter applied to the discretised system.
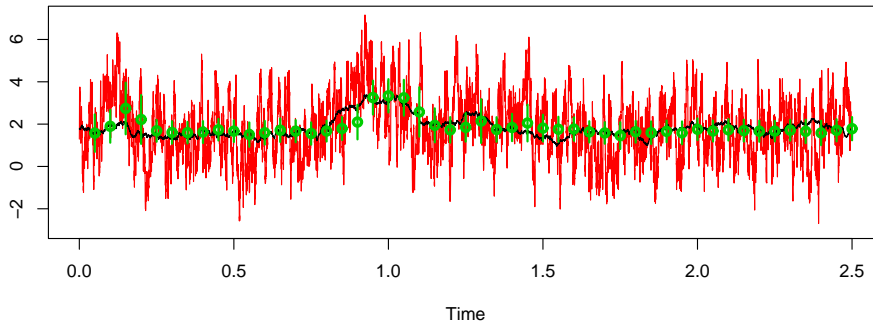
20

Figure 5: Results for the Fixman potential model. True state (black), observed process (red), filtered means (green circles) and regions plus/minus two standard errors (green lines). The RWPF was run for 500 filtering times; but for clarity results at only 50 are shown.

5 Extensions and future work

We have introduced importance and rejection sampling methods for solving the filtering problem for diffusions of the type in (1). As with any method based on importance or rejection sampling, the efficiency of our particle filters depends on how similar the proposal process is to the target. There are two approaches to improving the proposal process in our case. The first is by choosing an appropriate distribution from which to sample the pairs of particles $(x_{i-1}^{(j)}, x_i)$; and there is much work looking at this for standard particle filters (see Liu and Chen, 1998; Pitt and Shephard, 1999; Doucet *et al.*, 2000). The second is by choosing an appropriate law for the proposal of the path from time s_{i-1} to s_i, and is novel to the specific particle filter application we consider here. The work we have presented uses the law of the driftless version of the SDE (conditional on the start and end point, and the data); however a natural extension would be to consider the law of an SDE with linear drift (e.g. Shoji and Ozaki, 1998). This would have given an exact proposal in the OU example; and we would expect it to improve the efficiency of the methods for the other two examples we considered.

The importance sampling methodology we have presented can immediately be extended to the substantially larger family of models which arise by adding to (1) a linear term, i.e. models of the type

$$\mathrm{d}Z_s = DZ - \Sigma \nabla A(Z_s)\,\mathrm{d}s + \sqrt{2\Sigma}\mathrm{d}B_s\,, \quad s \in [0, T]\,, \quad Z_0 = z\,, \tag{13}$$

where D is a $d \times d$ matrix. A careful examination of our arguments reveal that in such a case we can modify the dominating process to be an Ornstein-Uhlenbeck process. Then

21

we obtain similar expressions as in Theorem 1; the Kalman smoother measure can still be easily simulated. Notice that this extension includes as a special case the problem where Y is not observed directly, but with linear additive error. Thus, the state of the system is (V, Y, X), where (Y, X) are evolving according to (1) and $dV_s = FV_s ds + dW_s$, where W is a Brownian motion (possibly dependent on the one that drives (Y, X)); we can easily include X in a linear fashion in the drift of V and include a constant diffusion coefficient. A simple calculation shows that the dynamics of (V, Y, X) are according to (13). In such a case we wish to reconstruct (Y, X) on the basis of discrete-time observations of V.

Another aspect of the simulation methodology of this paper is a connection with the representations of Del Moral and Miclo (2000b) which approximate the filtering distribution by means of certain interacting particle systems which evolve in continuous time. In such schemes particles die when a hazard function (similar to minus the exponent in (10)) exceeds an exponentially distributed time. Then, a randomly chosen existing particle duplicates. The rejection sampling methodology of Section 3.1 allows the exact simulation of the genealogy of such particle system.

An important future direction considers the case of multivariate diffusions without gradiant form drifts. Here the Itô formula argument to remove the stochastic integral in Girsanov's theorem (as in Theorem 1) does not apply. Instead, we require a methodology for estimating exponentials of stochastic integrals. Whilst this is far more challenging than the case considered in this paper, we are currently developing a theory of iterated forward and backward Itô integrals which provides a framework for stable estimation of these terms. A current goal of this programme of research therefore, is to translate this estimation theory into the general diffusion filtering problem context thus generalising substantially our contribution in this paper. This generalisation will allow our method to be applied to a wide range of applications other than the ones we consider here, such as signal processing and data assimilation.

Another interesting direction is to investigate the robustness of the proposed approach to the so-called micro-structure noise. Whilst diffusion models arise naturally when trying to find simple models which predict phenomena such as volatility, meta-stability and smooth empirical measures, the data are often incompatible with the diffusion assumption at small scales. For this reason, any inference procedure which uses high frequency data at small scales should be treated with caution. In the context of parameter estimation this issue is starting to be understood and examples of the issues arising, and strategies for dealing with them, see for example Ait-Sahalia *et al.* (2005b,a) in econometrics and Pavliotis and Stuart (2007); Pokern *et al.* (2008) in the physical sciences. In the context of filtering the issue is less well-understood.

Acknowledgements

Appendix

Appendix A: Notation

Let $A : R^d \to R$ denote a scalar field and $V : R^d \to R^n$ a vector field. The gradient operator ∇ applied to A yields a column vector with ith element $\partial A/\partial z_i$, $i = 1, \ldots, d$, and applied to V yields a $n \times d$ matrix with (i, j)th entry $\partial V_{ij}/\partial z_j$, $i = 1, \ldots, n$, $j = 1, \ldots, d$. Thus, $\nabla\nabla A$ is the Hessian matrix of A. For matrices L and K (of appropriate dimension) we will use the notation $L : K = \text{tr}(L^*K)$, where "tr" denotes the trace of a matrix and L^* the transpose of L. Then ":" is the inner-product on matrices which induces the Frobenius norm and $I : \nabla\nabla A$ is the Laplacian of A. We use $\langle \cdot, \cdot \rangle$ to denote the standard Euclidean inner-product on R^d,

Appendix B: Conditions on A

A natural way to get existence of solutions for (1), without imposing unnatural (for application) global Lipshitz conditions, is to work with a Lyapunov function. One set of conditions is:

(A0.1) $A \geq 0$ and level sets of A compact;

(A0.2) $\exists \alpha, \beta \geq 0 :$

$$-|\sqrt{\Sigma}\nabla A(z)|^2 + \Sigma : \nabla\nabla A(z) \leq \alpha - \beta A(z) \quad \forall z \in R^d.$$

Another possible condition is

(A0.1')
$$\exists \alpha, \beta \geq 0 : \langle \Sigma\nabla A(z), z \rangle \geq -\alpha + \beta|z|^2 \quad \forall z \in R^d.$$

Theorem 3. *Under either Assumptions (A0.1), (A0.2) or Assumption (A0.1') equation (1) has a unique solution for all $s \in [0, \infty)$.*

Proof. Under Assumptions (A0.1), (A0.2) define $V(z) = A(z)$ and under Assumptions (A0.1') define $V(z) = \frac{1}{2}|z|^2$. Notice that level sets of V are compact in both cases. By the Itô formula we have, in both cases,

$$V(Z_s) \leq V(Z_0) + \int_0^s \Big(\alpha - \beta V(Z_\tau)\Big)d\tau + M_s$$

23

where M_s is an Itô stochastic integral, and hence a Martingale. Taking expectations gives

$$\mathbb{E}V(Z_s) \leq \mathbb{E}(V(Z_s)) + \int_0^s \Big(\alpha - \beta\mathbb{E}V(Z_\tau)\Big)d\tau.$$

Applying the Gronwall lemma shows that $\mathbb{V}(Z_s)$ is bounded in terms of $\mathbb{V}(Z_0)$. Since level sets of V are bounded this prevents blow-ups and hence global existence of solutions follows. $\qquad\square$

Appendix C: Proof of Theorem 1

The mild regularity conditions refered to in the statement permit Girsanov's formula to hold. A particularly useful and weak set of conditions to ensure this are given in Rydberg (1997).

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 expression for this log-density is given by

$$-\frac{1}{2}\int_0^* \nabla A(Z_s)^* \mathrm{d}Z_s - \frac{1}{4}\int_0^* \nabla A(Z_s)^* \Sigma \nabla A(Z_s)\mathrm{d}s. \qquad (14)$$

Due to (A1) we can apply Itô's formula to eliminate the stochastic integral. (A2) allows us to define the positive function ϕ and re-write the log-density as in (3).

Appendix D: Densities of end-point biased measures

Let $W = (W_s, s \in [0, t])$ and $Z = (Z_s, s \in [0, t])$, for some $t > 0$, be two stochastic processes on R^d with corresponding measures $\mathbb{W}$ and $\mathbb{Z}$. Let $\nu_{0,t}$ and $\mu_{0,t}$ be the joint densities of their end-points (W_0, W_t) and (Z_0, Z_t) respectively with identical support. If it is true that the law of $(W \mid W_0 = x_0, W_t = x_1)$ is the same as the law of $(Z \mid Z_0 = x_0, Z_t = x_1)$, then the Radon-Nikodym derivative of $\mathbb{W}$ with respect to $\mathbb{Z}$ evaluated on an arbitrary path ω, is given by $\nu_{0,t}(\omega_0, \omega_t)/\mu_{0,t}(\omega_0, \omega_t)$. The proof of this result is along the lines of the proof of Proposition 1 of Beskos *et al.* (2006a) and it is omitted. In the context of Section 3.1 this result is applied to $\mathbb{W}(y_{0:s_i}, \pi_{s_{i-1}}^{(N)})$ and $\mathbb{W}(y_{0:s_i}, \nu_{s_{i-1},s_i}^{(N)})$ for $\nu_{s_0,s_1}^{(N)}$ the density proportional to (7).

Appendix E: Poisson expansion

Consider the following general problem. Let Φ be an unknown quantity, and we wish to obtain an estimator of $e^{-\Phi}$. Let $\tilde{\Phi}$ be an unbiased estimator of Φ, i.e. $\mathbb{E}[\tilde{\Phi} \mid \Phi] = \Phi$, and take $\tilde{\Phi}_j$ independent (conditionally on Φ) copies of $\tilde{\Phi}$. Notice that $\mathbb{E}[\prod_{j=1}^i (c - \tilde{\Phi}_j) \mid \Phi] = (c - \Phi)^i$. Then, for any $c \in R, \delta > 0$ we have:

$$e^{-\Phi} = e^{-c}\sum_{i=0}^\infty \left(\frac{c-\Phi}{\delta}\right)^i \delta^i/i! = e^{-c}\sum_{i=0}^\infty \mathbb{E}\left[\prod_{j=1}^i \frac{c-\tilde{\Phi}_j}{\delta} \mid \Phi, c, \delta\right]\delta^i/i! = e^{\delta-c}\mathbb{E}\left[\prod_{j=1}^\kappa \frac{c-\tilde{\Phi}_j}{\delta} \mid \Phi, c, \delta\right]$$

24

where $\kappa \sim \text{Poisson}(\delta)$, and the product $\prod_{j=1}^{0}$ is defined to be equal to 1. Thus, we have an unbiased estimator of $e^{-\Phi}$. If we fix c, it can be easily shown that the value of δ which minimises the variance of the estimator is given by $\mathbb{E}[(c - \tilde{\Phi})^2 \mid \Phi, c]^{1/2}$. Note that by taking $c = \delta$, and $\delta \to \infty$, the estimator converges almost surely to $e^{-\Phi}$.

In the context of Section 3.2, $\Phi = \int_{s_{i-1}}^{s_i} \phi(Y_s, X_s)\mathrm{d}s$, where (Y, X) is a path proposed according to $\mathbb{W}(y_{s_{i-1}, s_i}, x_{i-1}^{(j)}, x_i)$. Then $\tilde{\Phi} = (s_i - s_{i-1})\phi(Y_\psi, X_\psi)$, for ψ uniformly distributed on (s_{i-1}, s_i). Therefore, in order to obtain an estimator we only need to simulate a finite-dimensional distribution of $\mathbb{W}(y_{s_{i-1}, s_i}, x_{i-1}^{(j)}, x_i)$, at the Poisson times $\psi_1, \ldots, \psi_\kappa$, according to (W3). We have taken $c = r(s_i - s_{i-1})$, and chosen δ proportional to the time increment, as $\delta = \lambda(s_i - s_{i-1})$, for some $\lambda > 0$.

Appendix F: Implementation for Examples

We now give details of the approximate bounds we construct on $\phi(x, y)$ for the three simulated examples. Firstly we consider calculating approximate bounds on the X and Y processes under the proposal distribution.

Given x_{i-1} and x_i, the proposal for X is a Brownian Bridge, with marginally X_t have a normal distribution with mean $\mu(t) = x_{i-1} + (x_i - x_{i-1})(t - s_{i-1})/(s_i - s_{i-1})$ and variance $\sigma^2(t) = (t - s_{i-1})(s_i - t)/(s_i - s_{i-1})$. Thus our bound for the X process is given be the minimum and maximimum of $\{\mu(t) + c_X \sigma(t)\}$. This bound depends on c_X, and the probability that the X process goes outside the range given by this bound decays exponentially as $c_X \to \infty$. The value of c_X was chosen to be 2 for the OU and Fixman potential examples, and 3 for the double well potential examples.

As the Y process is observed at a finer scale, we use a simpler bound. Let δ denote the largest time-interval between successive observations, then we bound the Y process by $\min(y_{s_{i-1}:s_i}) - c_Y\sqrt{(\delta)}$ and $\max(y_{s_{i-1}:s_i}) + c_Y\sqrt{(\delta)}$. For the high-frequency data we simulated, we found $c_Y = 0$ to be sufficient.

Now we consider details of the three examples individually.

OU example

For this example

$$\phi(y, x) = (x - 0.9y)^2/2 + (y - 0.9x)^2/(2\epsilon).$$

The bound on ϕ was obtained from the bound on X and Y by substituting in the 4 corners of the bounding rectangle for X, Y; and choosing the largest resulting values of ϕ.

Double-Well example

For this example

$$\phi(y, x) = 32(y - x)^2 + 2(y^3 + 3y - 4x)^2/\epsilon - 3y^2/\epsilon.$$

The bound on ϕ was obtained from the bound on X and Y by substituting in the 4 corners of the bounding rectangle for X, Y; and choosing the largest resulting values of ϕ. Note

25

that in this case, this may not be a bound on ϕ for the given region of X, Y. Note that in this case the use of the stopping-time approach to dealing with negative weights is an advantage over truncation – the result of a poor bound on $\phi(x, y)$ at some time-steps will lead only to an increase in computation at the time-step and not to an introduction of large approximation error into the filter.

Fixman Potential example

For this example

$$\phi(y, x) = \frac{1}{4}\{[0.5(x - y) + 2\cos(2x) + 6\cos(3x)]^2 + [0.5(y - x)]^2/\epsilon + [8\sin(2x) + 36\sin(3x)]/\epsilon\}$$

Bounding ϕ for a given bound on (X, Y) is particularly difficult here due to the periodic behaviour of $\phi(x, y)$ as a function of x. A simple procedure which worked well in practice was to consider three values for x: the upper and lower bound on the X process, and the mean of the two bounds; together with two values for y: the upper and lower bound on the Y process. We evaluate $\phi(x, y)$ at the 6 combinations of x and y values and choose our bound to be the largest value of $\phi(x, y)$. As for the Double-Well example, the use of the stopping-time scheme to correct for negative weights increases the robustness of the filter to occasions were this procedure produces a poor bound.

References

Ait-Sahalia, Y., Mykland, P. and Zhang, L. (2005a). How often to sample a continuous-time process in the presence of market microstructure noise. *Rev. Financ. Studies* **18**, 351–416.

Ait-Sahalia, Y., Mykland, P. and Zhang, L. (2005b). A tale of two time scales: Determining integrated volatility with noisy high-frequency data. *J. Amer. Stat. Assoc.* **100**, 1394–1411.

Benes, V. E. (1981). Exact finite-dimensional filters for certain diffusions with nonlinear drift. *Stochastics* **5**(1-2), 65–92.

Bensoussan, A., Lions, J.-L. and Papanicolaou, G. (1978). *Asymptotic analysis for periodic structures*, volume 5 of *Studies in Mathematics and its Applications*. North-Holland Publishing Co., Amsterdam.

Beskos, A., Papaspiliopoulos, O. and Roberts, G. O. (2005). A new factorisation of diffusion measure and finite sample path constructions. Submitted.

Beskos, A., Papaspiliopoulos, O. and Roberts, G. O. (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* .

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

Cioranescu, D. and Donato, P. (1999). *An introduction to homogenization*, volume 17 of *Oxford Lecture Series in Mathematics and its Applications*. The Clarendon Press Oxford University Press, New York.

Crisan, D. and Lyons, T. J. (1999). A particle approximation of the solution of the kushner-stratonovitch equation. *Probability Theory and Related Fields* **115**, 549–578.

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.

Del Moral, P. and Miclo, L. (2000a). Branching and interacting particle systems approximations of Feynman-Kac formulae with applications to non-linear filtering. In: *Séminaire de Probabilités, XXXIV*, volume 1729 of *Lecture Notes in Math.*, Springer, Berlin, 1–145.

Del Moral, P. and Miclo, L. (2000b). A Moran particle system approximation of Feynman-Kac formulae. *Stochastic Process. Appl.* **86**(2), 193–216.

Del Moral, P., Jacod, J. and Protter, P. (2001). The Monte-Carlo method for filtering with discrete-time observations. *Probab. Theory Related Fields* **120**(3), 346–368.

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

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., Springer, New York, 3–14.

Fearnhead, P., Papaspiliopoulos, O. and Roberts, G. O. (2006). Particle filters for partially observed diffusions. In revision for J. Roy. Statist. Soc. Ser. B.

Ferrante, M. and Vidoni, P. (1998). Finite-dimensional filters for nonlinear stochastic difference equations with multiplicative noises. *Stochastic Process. Appl.* **77**(1), 69–81.

Fixman, M. (1978). *J. Chem. Phys.* **69**, 1527–1538.

Gardiner, C. W. (1985). *Handbook of stochastic methods*, volume 13 of *Springer Series in Synergetics*. Springer-Verlag, Berlin, 2nd edition, for physics, chemistry and the natural sciences.

Genon-Catalot, V. (2003). A non-linear explicit filter. *Statist. Probab. Lett.* **61**(2), 145–154.

Givon, D., Kevrekidis, I. G. and Kupferman, R. (2006). Strong convergence of projective integration schemes for singularly perturbed stochastic differential systems. *Commun. Math. Sci.* **4**(4), 707–729.

Jourdain, B. and Sbai, M. (2007). Exact retrospective Monte Carlo computation of arithmetic average Asian options. Available from https://hal.archives-ouvertes.fr/hal-00141141.

Kalman, R. and Bucy, R. (1961). New results in linear filtering and prediction theory. *Journal of Basic Engineering, Transacation ASME series D* **83**, 95–108.

Liu, J. S. and Chen, R. (1998). Sequential Monte Carlo methods for dynamic systems. *J. Amer. Statist. Assoc.* **93**(443), 1032–1044.

Metzner, P., Schütte, C. and Vanden-Eijnden, E. (2006). Illustration of transition path theory on a collection of simple examples. *The Journal of Chemical Physics* **125**, 084110.

Papavasiliou, A. and Kevrekidis, I. (2007). Variance reduction for the equation-free simulation of multiscale stochastic systems. *Multiscale Model. Simul.* **6**(1), 70–89.

Pavliotis, G. and Stuart, A. (2007). Parameter estimation for multiscale diffusions. *J. Stat. Phys.* **127**(4), 741–781.

Pavliotis, G. and Stuart, A. (2008). *Multiscale Methods: Averaging and Homogenization*. Springer, current version availalbe from http://www.maths.warwick.ac.uk/ stuart/book.pdf.

Peluchetti, S. (2007). *Aspects of the efficiency of the exact algorithm for diffusion simulation*. Ph.D. thesis, Boconni University, Milan.

Pitt, M. K. and Shephard, N. (1999). Filtering via simulation: auxiliary particle filters. *J. Amer. Statist. Assoc.* **94**(446), 590–599.

Pokern, Y., Pavliotis, G. and Stuart, A. (2008). Parameter estimation for multiscale diffusions: An overview. In preparation.

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

Shoji, I. and Ozaki, T. (1998). Estimation for nonlinear stochastic differential equations by a local linearization method. *Stochastic Analysis and Applications* **16**.

Siegmund, D. (1985). *Sequential Analysis*. Springer-Verlag.

Vanden-Eijnden, E. (2003). Numerical techniques for multi-scale dynamical systems with stochastic effects. *Commun. Math. Sci.* **1**(2), 385–391.