Inference and Filtering for Partially Observed Diffusion Processes via Sequential Monte Carlo

E. L. Ionides

University of Michigan, Department of Statistics
Technical Report # 405
September 9, 2003

Abstract

Diffusion processes observed partially or discretely, possibly with observation error, arise when constructing stochastic models in continuous time. The method of Sequential Monte Carlo provides an alternative to Markov Chain Monte Carlo methods, and can be effective in complex models at the cutting edge of scientific research. This paper introduces Sequential Monte Carlo approaches to inference for partially observed diffusion processes.

New methods for solving the filtering, predicting and smoothing problems are developed. Two new filtering algorithms are compared with existing methods on a nonlinear problem for which a closed form solution exists. A novel measure of filter accuracy helps to highlight strengths and weaknesses of the methods.

1 Introduction

When trying to simulate a realization $x_{1,T} = (x_1, \ldots, x_T)'$ from a high dimensional density $f_T(x_{1,T})$ it may help to work sequentially, drawing $x_{1,t}$ from some density $f_t(x_{1,t})$ then updating $x_{1,t}$ to generate a draw from $f_{t+1}(x_{1,t+1})$. This approach has been used for computer simulation of long-chain polymers since the 1950's (Hammersley & Morton 1954, Rosenbluth & Rosenbluth 1955). We refer to such a technique as Sequential Monte Carlo (SMC). Many researchers independently discovered that SMC is readily applicable to state space models (Gordon, Salmond & Smith 1993, Kitagawa

1996, Isard & Blake 1996, Del Moral 1996, Crisan & Lyons 1997). A state space model arises when x_t is a Markov process, called the state process, and there is an observed process y_t with conditional densities $f(y_t|y_{1,t-1}, x_{1,t}) = f(y_t|x_t)$. One typically wishes to sample from the conditional densities $f(x_t|y_{1,t})$ [filtering], $f(x_{t+1}|y_{1,t})$ [prediction] or $f(x_t|y_{1,T})$ [smoothing]. We assume these densities exist, and note the convention that $f(\cdot \mid \cdot)$ is a generic density which is then specified by its arguments. The densities may further depend on an unknown parameter vector θ for which inference is required. State space models have many applications including signal processing (Kalman 1960, Anderson & Moore 1979), time series analysis (Shumway & Stoffer 2000, Durbin & Koopman 2001), economics (Harvey 1989), finance (Shephard & Pitt 1997), biology (Ionides, Fang, Isseroff & Oster 2003), meteorology (Evensen & van Leeuwen 1996), neuroscience (Brown, Frank, Tang, Quirk & Wilson 1998), and various others (Doucet, de Freitas & Gordon 2001, Section IV). It appears that real world phenomena are often well modeled by some Markov process with sufficiently rich state space, perhaps constructed according to physical or chemical or economic principles, about which we can make only noisy or incomplete observations.

Here, we are concerned with the situation in which $x_{1,T}$ is a discretization of a continuous time process $\{x_t, 1 \leq t \leq T\}$, with continuous sample paths. This includes all the applications listed above, while ruling out inherently discrete state space models such as genetic inheritance (Lander & Green 1987) and speech recognition (Rabiner 1989). Let us take $\{x_t\}$ to be an Ito diffusion in $\mathbb{R}^m$,

$$dx_t = \mu(x_t, t, \theta)dt + \sigma(x_t, t, \theta)dW_t, \tag{1}$$

with σ an $m \times m$ matrix and $\{W_t\}$ Brownian motion in $\mathbb{R}^m$. We assume that μ and σ are sufficiently regular to guarantee the existence of a unique solution to (1) (Oksendal 1998, Theorem 5.2.1). Suppose y_t lies in $\mathbb{R}^n$ and is given by

$$y_t = C(t, \theta)x_t + D(t, \theta) + \eta_t, \tag{2}$$

with $\eta_t \sim N(0, \tau(t, \theta)\tau^1(t, \theta))$. For now, we assume that $\sigma\sigma'$ and $\tau\tau'$ are invertible. The dependence of μ, σ, C and τ on t and θ will only be noted explicitly where required. The methods of this paper also apply to a general observation equation,

$$y_t = g\left(x_t, \eta_t, \theta, t\right),$$

with a linear/Gaussian approximation of the form (2).

Two new filtering algorithms appropriate to the system (1) and (2) are proposed in Sections 1.2 and 2.5. Section 2 discusses issues arising in the implementation of

these algorithms. Section 3 works through an example. Filters are compared by their ability to approximate the log likelihood of the observations, and this criterion is examined in Section 4. Section 5 investigates the situation when $\tau\tau'$ is not invertible. Section 6 is a concluding discussion.

1.1 A Shortcoming of Standard SMC Methods for Partially Observed Diffusions

A standard SMC approach, sometimes termed the particle filter, is to solve (in a sequential Monte Carlo sense) the prediction equation

$$f(x_{t+1}|y_{1,t}) = \int f(x_t|y_{1,t}) f(x_{t+1}|x_t) dx_t \tag{3}$$

and the filtering equation

$$f(x_t|y_{1,t}) = \frac{f(x_t|y_{1,t-1}) f(y_t|x_t)}{\int f(x_t|y_{1,t-1}) f(y_t|x_t) dx_t}. \tag{4}$$

The solution is recursive. Suppose $\{X^F_{t-1,j}, j = 1, \ldots, J\}$ has an empirical distribution approximating $f(x_{t-1}|y_{1,t-1})$. Then using (3), each "particle" $X^F_{t-1,j}$ can be moved according to the transition density of the state space to give a particle $X^P_{t,j}$ whose marginal distribution approximates the prediction density, $f(x_{t+1}|y_{1,t})$. The filtering problem is then solved by setting $\{X^F_{t,j}\}$ to be a sample drawn with replacement from $\{X^P_{t,j}\}$ with sampling weights given by

$$w_j = f(y_t|x_t{=}X^P_{t,j}).$$

This method, and many variations proposed in (Doucet, de Freitas & Gordon 2001), have the difficulty that they work poorly when the observation error is small. Heuristically, in this situation $\{w_j\}$ should have high variance since most of the proposal steps will be inconsistent with the new observation. In the extreme case, where the density $f(x_{t+1}|y_{1,t+1})$ exists on a region of lower dimension than $f(x_{t+1}|y_{1,t})$, all the w_j are zero and the algorithm breaks down. This will typically happen if $\tau\tau'$ is not invertible.

In the case where $\tau = 0$ and $C = I$, the identity matrix, (1) and (2) specify a discretely observed diffusion process, which has applications such as economics and finance. Much work has been done in this case (see Elerian, Chib & Shephard (2001), and references therein) but the framework considered here provides some fresh insights and new methodology. This situation is given special consideration in Section 5.

1.2 A New SMC Approach - the Conditional Particle Filter (CPF)

It has been pointed out by Liu, Chen & Logvinenko (2001) that the prediction step can be replaced by a proposal step in a general SMC framework. Pitt & Shepard (2001) investigated making the proposal distribution depend on the observation process, and coined the term "adaptive particle filter" for such an algorithm. Our approach to remedy the difficulties discussed in Section 1.1 is to develop an adaptive particle filter which is convenient for systems such as (1) and (2). (3) and (4) may be replaced by a proposal step from a density

$$f_p(x_t, x_{t+1}|y_{1,t+1}) = f(x_t|y_{1,t})f(x_{t+1}|x_t, y_{t+1}), \tag{5}$$

followed by a filtering step

$$f(x_{t+1}|y_{1,t+1}) \propto \int f_p(x_t, x_{t+1}|y_{1,t})f(y_{t+1}|x_t)dx_t. \tag{6}$$

This works because

$$\begin{aligned} f(x_{t+1}|y_{1,t+1}) &= \int f(x_t, x_{t+1}|y_{1,t+1})dx_t \\ &= \int f(x_t|y_{1,t+1})f(x_{t+1}|x_t, y_{t+1})dx_t \\ &= \int \frac{f(x_t|y_{1,t})f(y_{t+1}|x_t)}{f(y_{t+1}|y_{1,t})} f(x_{t+1}|x_t, y_{t+1})dx_t. \end{aligned}$$

The proposal step is solved (in a sequential Monte Carlo sense) by moving each particle $X^F_{t-1,j}$ to $X^P_{t,j}$ by drawing from $f(x_t|y_t, x_{t-1}{=}X^F_{t-1})$. The filtering step then assigns weight w_i to $X^P_{t,j}$ given by

$$w_j = f(y_t|x_{t-1}{=}X^F_{t-1,j}). \tag{7}$$

$\{X^F_{t,j}\}$ can then be constructed by resampling with replacement from $\{X^P_{t,j}\}$ with weights w_j. The key to the algorithm is that, for the class of models including (1) and (2), this preliminary trial and filtering steps can be solved (again, in a Monte Carlo sense). We call this algorithm a conditional particle filter (CPF). Section 2 gives the details necessary to implement a CPF algorithm.

2 Implementing Conditional Particle Filters

Section 2.1 discusses simulating from $f(x_t|x_{t-1}, y_t)$ and Section 2.2 discusses estimating $f(y_t|x_{t-1})$. All SMC methods usually involve resampling, and Section 2.3 discusses resampling issues for CPF. Section 2.4 discusses numerical solutions to SDEs. Carrying out CPF using a crude numerical solution gives rise to an algorithm of interest in its own right, presented in Section 2.5, which we call a particle Kalman filter (PKF). Parameter estimation and possible extensions are mentioned in Sections 2.6 and 2.7.

2.1 Simulating Conditional Diffusions

Conditioning the diffusion $\{x_t\}$ on x_o and y_1 results in a conditional diffusion $\{\hat{x}_t\} = \{x_t|x_0, y_1\}$. The SDE for $\hat{x}_t$ is $d\hat{x}_t = \hat{\mu}(\hat{x}_t)dt + \sigma(\hat{x}_t)dW_t$, for $0 \leq t \leq 1$. Here $\sigma(\hat{x}_t)$ is the same as in (1) and $\hat{\mu}(\hat{x}_t)$ is not in general easy to calculate. Nevertheless, methods building on Roberts & Stramer (2001) and Ozaki (1992) can be used to simulate $\{\hat{x}_t, 0 \leq t \leq 1\}$ based on a local linearization.

Let $\tilde{x}_t$ be an approximation to $\hat{x}_t$, with SDE

$$d\tilde{x}_t = \tilde{\mu}(\tilde{x}_t)dt + \sigma(\tilde{x}_t)dW_t$$

Write $P, \hat{P}$ and $\tilde{P}$ for the laws of $\{x_t \mid x_0\}$, $\{\hat{x}_t\}$ and $\{\tilde{x}_t\}$ respectively, for $0 \leq t \leq 1$. A key point is that, to carry out importance sampling from $\hat{P}$ using trial distribution $\tilde{P}$ it is sufficient to be able to calculate $d\tilde{P}/dP$, which is proportional to $d\tilde{P}/d\hat{P}$. Thus, although $\tilde{\mu}$ should be similar to $\hat{\mu}$ in order to have a computationally efficient simulation, it is not necessary to know $\hat{\mu}$. This property is the functional form of the familiar identity

$$f(x|y) = \frac{f(x,y)}{f(y)} \propto f(x,y),$$

where the constant of proportionality depends on y. The case where $\tau\tau'$ is singular requires more care, as $d\hat{P}/dP$ may not exist, and is dealt with in Section 5.

2.2 Estimating the Conditional Density

An analogous situation to calculating the likelihood in a state space model via the methods of Section 1.2, for a bivariate random variable (X, Y) with density $f(x, y)$, is to estimate $f(y)$ using

$$f(y) \approx \hat{f}(y) = \frac{1}{K}\sum_{k=1}^{K} f(y|x{=}X_k) \tag{8}$$

where X_k is a sample from density $f(x)$. If X_k is instead drawn from another density $g(x)$ we have an estimate

$$f(y) \approx \hat{f}(y) = \frac{1}{K} \sum_{k=1}^{K} \frac{f(X_k)}{g(X_k)}. \tag{9}$$

The density $g(x)$ can be chosen to depend on the observation, y, and so the ideal choice would be $g(x) = f(x|y)$, in which case $\mathrm{Var}(\hat{f}(y)) = 0$. Of course, $f(x|y)$ is not a density that can usually be calculated explicitly for nonlinear state space models though it can be sampled from using the methods of Section 2.1. Instead, we take $g(x)$ to be a linearized approximation to $f(x|y)$. Using the notation of Section 2.1, we can rewrite (9) in the form required in (7), as

$$\begin{aligned} f(y_1|x_0) &= E_P[f(y_1|x_1)] \\ &= E_{\tilde{P}}[f(y_1|x_1{=}\tilde{x}_1)\frac{dP}{d\tilde{P}}] \\ &\approx \frac{1}{K}\sum_{k=1}^{K} f(y_1|x_1{=}\tilde{x}_1^{(k)})\frac{dP}{d\tilde{P}}\{\tilde{x}_t^{(k)}\} \end{aligned} \tag{10}$$

where $\{\tilde{x}_t^{(k)}\}$ is a sample from $\tilde{P}$. Conveniently, the same sample from $\tilde{P}$ can be used to compute (10) as is used for importance sampling from $\{\hat{x}_t\}$ in Section 2.1.

2.3 Weights and Resampling

Manipulating weights in importance sampling type methods, such as SMC, can be done following the principle of properly weighted samples (Liu, Chen & Logvinenko 2001).

Definition A weighted sample $\{(X_i, w_i), i = 1, \ldots, N\}$ is *properly weighted* for a density $f(x)$ if, for some constant C and any square integrable function $h(x)$,

$$E[h(X_i)w_i] = C \int h(x)f(x)dx.$$

For any set of positive resampling weights $\{a_i\}$, sampling with replacement from $\{(X_i, w_i/a_i)\}$ with weights a_i results in a new (approximately) proper sample for $f(x)$ (Rubin 1987). The originally used SMC resampling weights were $a_i = w_i$ (Kitagawa

1996, Gordon, Salmond & Smith 1993). These weights prune heavily any X_i with small weight w_i, allowing computation at future steps to concentrate on "successful" particles. In practice, lighter pruning, say with $a_i = \sqrt{w_i}$, may be preferable (Liu, Chen & Logvinenko 2001).

In fact, for the CPF and PKF algorithms proposed here, resampling may be unnecessary. Heuristically, the reason for this is that for these algorithms the particles can be expected to have mixing properties. In the original SMC each particle $\tilde{x}_t$ moves around the state space according to the unconditional state Markov transition density $f(x_t|x_{t-1})$. Pruning is then required to remove particles whose location is unlikely given the observation process. In CPF and PKF, a particle $\hat{x}_t$ moves according to a conditional Markov transition density $f(x_t|x_{t-1}, y_t)$. Mixing properties of similar conditional processes have been studied by (Bickel & Ritov 1996, Bickel, Ritov & Ryden 1998, Jensen & Petersen 1999). If mixing occurs, then the distribution of $\hat{x}_t$ should not stray too far from the filtering distribution $f(x_t|y_{1,t})$. The proper weight of $\hat{x}_t$ for $f(x_t|y_{1,t})$, the *filtering weight*, is

$$w_t^{(f)} = \prod_{s=1}^{t} v_s$$

where $v_s = f(y_s|x_{s-1} = \hat{x}_{s-1})$. When t is not small, $w_t^{(f)}$ typically has high variance. However, if $\hat{x}_t$ is mixing then the recent weights will be more relevant, so truncated weights

$$w_{t,k}^{(f)} = \prod_{s=t-k}^{t} v_s$$

can be used. A similar issue arises in smoothing, where $(\hat{x}_t, w_t^{(s)})$ is properly weighted for $f(x_t|y_{1,T})$ with *smoothing weight* and truncated smoothing weight given by

$$w_t^{(s)} = \textstyle\prod_{s=1}^{T} v_s, \quad w_{t,k}^{(s)} = \prod_{s=t-k}^{t+k} v_s.$$

Although these truncated weights somewhat resemble previous methods using truncated smoothers such as (Anderson & Moore 1979, Sections 7.3 and 9.6) and (Kitagawa & Sato 2001), this particular method appears new. Note it is only the weights and not the particle process that is truncated. A reasonable value of k can be selected by bias/variance tradeoff considerations. A large k increases the Monte Carlo variance of an estimated quantity (such as the mean or quantiles of the conditional state process, or the likelihood of the observations). A small k introduces bias due to the truncation.

Note that there may be room for improvement on truncation as a way of taking advantage of mixing. Filtering weights of the form

$$w_t^{(f)} = \exp\Big\{ \sum_{s=1}^{t} \rho^{t-s} \log v_s \Big\}$$

might be considered, for example.

2.4 Numerical Methods

For a practical implementation of CPF the necessary sample paths and stochastic integrals must be discretely approximated. Appropriate numerical methods are discussed by Kloeden & Platen (1999). Different algorithms arise for weak and strong approximations. Here, we use strong approximations in order to guarantee the correct joint distributions of solutions to varying SDEs driven by the same Brownian motion.

The Euler scheme, or order 0.5 strong Taylor approximation, to

$$dx_t = \mu(x_t, t)dt + \sigma(x_t, t)dW_t$$

is given by

$$x_{t+\delta} = x_t + \mu\delta + \sigma\Delta W_t$$

where $\Delta W_t = W_{t+\delta} - W_t$. If P is the law of $\{x_t, 0 \le t \le N\delta\}$ and $\tilde{P}$ is the law of the solution $\{\tilde{x}_t, 0 \le t \le N\delta\}$ to

$$d\tilde{x}_t = \tilde{\mu}(\tilde{x}_t, t)dt + \sigma(\tilde{x}_t, t)dW_t$$

then the likelihood ratio evaluated at $\{\xi_t, 0 \le t \le N\delta\}$, denoted as $\frac{d\tilde{P}}{dP}\{\xi_t\}$, can be approximated by

$$\exp\left\{ \frac{1}{2\delta} \sum_{n=0}^{N-1} [2\left(\tilde{\mu}_{n\delta} - \mu_{n\delta}\right)' \left(\sigma\sigma'\right)^{-1} \Delta\xi_{n\delta} + \mu'_{n\delta}(\sigma\sigma')^{-1}\mu_{n\delta} - \tilde{\mu}'_{n\delta}(\sigma\sigma')^{-1}\tilde{\mu}_{n\delta}] \right\}. \quad (11)$$

Note that (11) is both a Riemann sum approximation to the integral given by the Girsanov theorem and the exact likelihood ration of the numerical approximation. For higher order Taylor approximations, these two quantities will usually be different. The exact likelihood ratio of the approximation appears more reasonable to use, though this would be hard to find for more complex schemes. The exception to this is the

Milstein scheme (order 1.0 strong Taylor approximation) in the special case where W_t and x_t are in $\mathbb{R}^m$ with $m = 1$, or when the noise is commutative (Kloeden & Platen 1999, Section 10.3). In the case $m = 1$, the Milstein scheme is

$$x_{t+\delta} = x_t + \mu\delta + \sigma\Delta W_t + \frac{1}{2}\sigma\frac{\partial\sigma}{\partial x}\left((\Delta W_t)^2 - \delta\right).$$

2.5 The Particle Kalman Filter (PKF)

The Euler scheme of Section 2.4 with $N = \delta = 1$ corresponds to a local Gaussian approximation which may be written as

$$x_t|x_{t-1} \sim N\left(A(x_{t-1}), \Phi(x_{t-1})\right). \tag{12}$$

A corresponding local Gaussian observation equation is

$$y_t|x_t, x_{t-1} \sim N\left(C(x_{t-1})x_t + D(x_{t-1}), \Psi(x_{t-1})\right). \tag{13}$$

The PKF consists of applying the recursions in (5) and (6) to the system specified by (12) and (13). The model given by (12) and (13) is similar to the linearization employed in the Extended Kalman filter (EKF), the widely used and fairly successful technique of linearizing a nonlinear model and then applying the Kalman filter (Anderson & Moore 1979). The EKF however is not an exact filter, even for the linearized model, in the sense that it does not find the true conditional distribution of the state process. In particular, the EKF estimate of the conditional distribution is necessarily unimodal. The PKF, on the other hand, gives an exact solution, up to Monte Carlo error, of the locally Gaussian system of (12) and (13). This system can be viewed as a discretization of a diffusion process, but also has some interest in its own right. In phenomenological modeling, e.g. of financial or physical systems, the underlying continuous time model may be a heuristic that can be discarded once it has been used to motivate a discrete time model.

The observation equation (13) has the interesting feature that C, D and Ψ can depend on x_{t-1}. This useful fact arises from (5) and (6), and has the consequence that the filtering weights are $f(y_t|x_{t-1})$ rather than the usual $f(y_t|x_t)$. Also, PKF does not require a numerical integration over x_t to find $f(y_t|x_{t-1})$, since the conditional Gaussian distribution is known. In addition, $f(y_t|x_t)$ may be poorly behaved, when some linear combination of components of x_t can be observed with little or no error, whereas $f(y_t|x_{t-1})$ has extra state space variation which may push the distribution away from singularity. The author is not aware of previous work on methods closely resembling PKF, although it appears a simple combination of particle filtering and (extended) Kalman filtering.

2.6 Parameter Estimation

The likelihood may be calculated via

$$f(y_{1,T}) = \prod_{t=1}^{T} f(y_t|y_{1,t-1})$$

using a Monte Carlo approximation, such as CPF or PKF, to the integral

$$f(y_t|y_{1,t-1}) = \int f(y_t|x_{t-1}) f(x_{t-1}|y_{1,t-1}) dx_{t-1}.$$

For CPF, $f(y_t|x_{t-1})$ can be estimated by Monte Carlo methods, following the discussion in Section 2.2. For PKF this quantity can be found analytically. Remembering that there is a parameter, θ, implicit in all the model equations, inference for θ may be carried out using the usual Bayesian or frequentist methods.

When no resampling is carried out (see Section 2.3), the Monte Carlo variation can be dealt with by fixing the seed of the random number generator, finding estimates $\hat{\theta}_1, \ldots, \hat{\theta}_L$ at different seed values, and combining these estimates.

In the presence of resampling, the branching of the particles depends on θ. One can simply sample the likelihood with Monte Carlo error at different parameter values or adopt an importance sampling strategy such as Hürzeler & Künsch (2001). Hürzeler & Künsch (2001) also propose Monte Carlo EM algorithms, based on solving the smoothing problem.

Parameter estimation may be carried out by adding θ to the state space (Anderson & Moore 1979). This method is more successful if θ is a dynamic parameter, i.e., $\theta = \theta_t$ where θ_t is itself a diffusion process. If θ does not change with time, but $\theta = \theta_0$ has some prior distribution $f(\theta_0)$, then the filtering estimate $f(\theta|y_{1,t})$ is formally a Bayesian posterior distribution. However, as pointed out in Section 2.3, SMC works poorly when the conditional process $\hat{\theta}_t$ with transition density $f(\theta_t|\theta_{t-1} = \hat{\theta}_{t-1}, y_t)$ is slowly mixing. In the fixed parameter case, $\hat{\theta}_t = \theta_0$ so $\hat{\theta}_t$ does not "mix" at all. The θ component of the particles are then stuck at their initial positions, and only their weights change with time. This problem also arises with other particle based nonlinear filtering algorithms.

2.7 Extensions

The CPF method applies more or less immediately to certain generalizations of (1) and (2). If $\sigma\sigma'$ is not invertible, or similarly if $\{W_t\}$ lies in $\mathbb{R}^d$ and $\{x_t\}$ lies in $\mathbb{R}^m$ with $m \neq d$, then one can use a generalized inverse.

The observation equation (2) can be generalized to

$$y_t = g\left(\{x_s, t-1 \leq s \leq t\}, \eta_t, \theta, t\right),$$

where $\{\eta_t, t = 1, \ldots, T\}$ are independent random variables with density from a known parametric family. The requirement on g is that one can carry out a local linearization, as in the example of Section 3, by making a linear, Gaussian approximation to η_t and y_t. The closer this approximation , the more computationally efficient the Monte Carlo scheme will be. The observation process does not have to be real valued, it could be categorical or integer valued.

There is some interest in removing the requirement that x_t has continuous sample paths, by allowing dW_t in (1) to be the increment process for a Lévy process rather than Brownian motion (Barndorff-Nielsen, Kosch & Resnick 2001). It is not clear how useful the diffusion approximations developed here would be in that case.

Another possible extension, similar to that employed in the conditional Gaussian filter (Anderson & Moore 1979), is that the state evolution equation and observation equation can depend on previous observations.

3 An Example

For a simple numerical example, we consider a one dimensional nonlinear filtering problem for which an exact closed form filter exists, namely

$$dx_t = [-\alpha h^{-1}(x_t)h'(h^{-1}(x_t)) + \frac{1}{2}h''(h^{-1}(x_t))]dt + h'(h^{-1}(x_t))dW_t \tag{14}$$

$$y_t = h^{-1}(x_t) + \eta_t, \quad \eta_t \sim N(0, \tau^2) \tag{15}$$

where $h(\cdot)$ is an increasing function, and $h'(x) = dh/dx$. In this section only, h' does not represent the transpose of h. If $x_t = h(z_t)$ then z_t and y_t form a linear Gaussian system

$$dz_t = -\alpha z_t dt + dW_t \tag{16}$$

$$y_t = z_t + \eta_t \tag{17}$$

The likelihood of $\{y_t, t = 1, \ldots, T\}$ can be found using the Kalman filter, if the linear representation (16) is known. We try to approximate this likelihood using (14) and (15) with an extended Kalman filter (EKF), particle filter (PF), particle Kalman filter (PKF) and conditional particle filter (CPF).

3.1 Linearization for EKF and PKF

The linearization of the SDE

$$dx_t \;=\; \mu(x_t)dt + \sigma(x_t)dW_t$$

around the filter estimate x_s^f, $s \le t$, is

$$dx_t \;\doteq\; [\mu(x_s^f) + \mu'(x_s^f)(x_t - x_t^f)]dt + \sigma(x_t^f)dW_t$$

We use this linearization together with an Euler approximation to give a linear Gaussian discrete time system approximating (14) and (15), namely

$$\begin{array}{rcll} x_{t+1} &=& A_t x_t + B_t + \epsilon_t, & \epsilon_t \sim N[0, \Phi_t^2], \\ y_{t+1} &=& C_{t+1}x_{t+1} + D_{t+1} + \eta_{t+1}, & \eta_{t+1} \sim N[0, \tau^2]. \end{array} \tag{18}$$

The parameters A_t, B_t, Φ_t, C_{t+1} and D_{t+1} are functions of the filtering estimate x_t^f, the prediction estimate x_{t+1}^p, and also h, h', h'', h'''. The expressions are derived in Appendix C.

3.2 The Choice of Transformation

Here we consider the transformation

$$x = h(z) = [(|z| + 1)^2 - 1]sgn(z)$$

where $sgn(z) = \{1 \text{ if } z > 0, -1 \text{ if } z < 0\}$. Then

$$\begin{array}{rcl} h^{-1}(x) &=& [\sqrt{|x|} - 1]sgn(x) \\ h'(z) &=& 2(|z| + 1) \\ h''(z) &=& 2sgn(z). \end{array}$$

This choice of transform leads to long tailed distributions and to a peak in the conditional density $f(x_t \mid y_{1,t})$ at $x_t = 0$ (see Fig. 1). Using the inverse of this transform would give short tails and a bimodal stationary distribution for x_t.

3.3 Results

Table 1 compares four filters, using the accuracy measure developed in Section 4. It shows that CPF is the most accurate filter in this situation, followed by PF, KPF

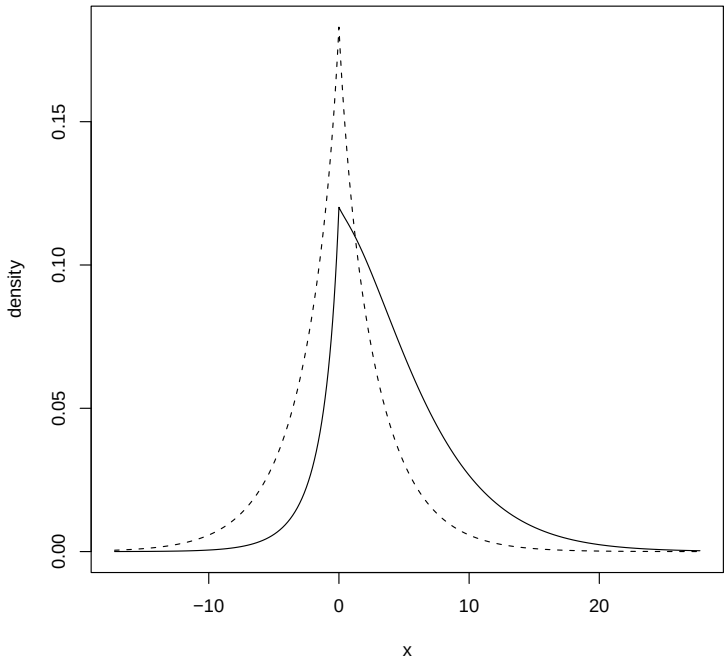

Figure 1: Prediction distribution for (14) and (15) when the prediction median is $x_t^p = 3$ (solid line) and $x_t^p = 0$ (dashed line), with $\tau = 1$ for large t.

Parameter	Description	Required for	Value
N_p	# of Particles	PF, KPF, CPF	200
N_t	# of Steps per Observation for Numerical Solution of SDE	PF, CPF	5
N_f	# of lags used for filtering (Sec. 2.3)	KPF, CPF	3
N_r	# of Trials for Importance Resampling	CPF	10

	τ	$b \times 10^2$	$\sqrt{V} \times 10^2$	$\sqrt{M} \times 10^2$	$A \times 10^2$
EKF	1	-3.32	25.1	N/A	25.4
KPF	1	-1.62	15.5	4.67	16.4
PF	1	-0.45	3.26	8.54	9.27
CPF	1	-0.16	3.94	3.99	5.67
EKF	0.25	-8.1	42.7	N/A	43.9
KPF	0.25	-5.3	30.7	2.94	31.8
PF	0.25	-2.52	13.7	21.7	29.8
CPF	0.25	-0.82	9.78	3.56	10.5

Table 1: Bias, b, filter error, $\sqrt{V}$, Monte Carlo error, $\sqrt{M}$, and accuracy, A. Calculated by simulation for (14) and (15), with parameter settings given above. Simulation errors are in the last nonzero digit presented.

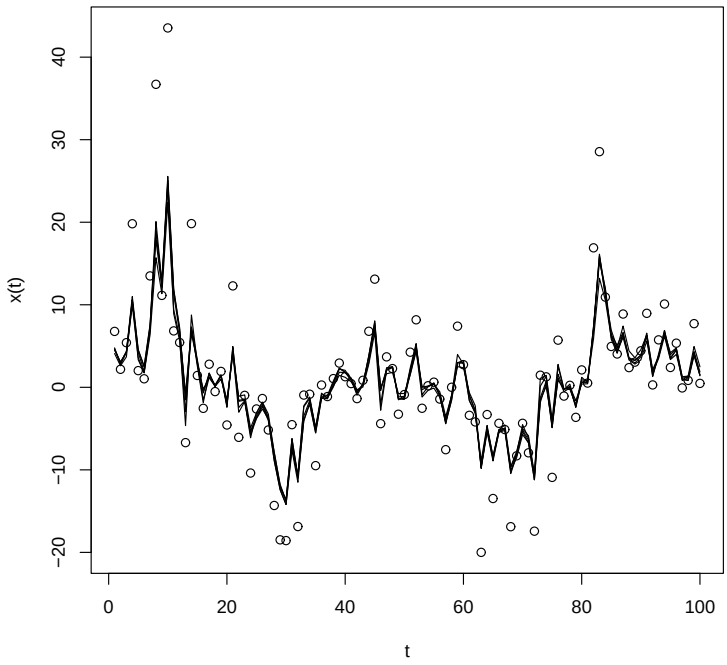

Figure 2: A realization from (14) and (15), with $T = 100$. Transformed observations, $h(y_t)$, are shown as points. The filtering means, estimating the unobserved process x_t, are shown superimposed for EKF, PF, PKF and CPF. The transformed Kalman filter for estimating z_t is also shown. All the methods agree closely on their point estimates of the unobserved process.

and finally EKF. The filter variance of EKF is large, because the conditional state distribution is not well approximated by a Gaussian (see Fig. 1). The particle filter, PF, has a relatively high Monte Carlo error, $\sqrt{M}$, especially when the observation error, τ, becomes small. On the other hand, $\sqrt{M}$ decreases with smaller τ for KPF and CPF. The filter error, $\sqrt{V}$, which is interpreted as the error after averaging over many replications of a Monte Carlo filter, is larger for each filter when $\tau = 0.25$. Techniques introduced in Section 5 for $\tau = 0$ could be used to reduce the loss of accuracy of CPF for small τ.

4 Evaluating Filters

In applications, the main goal of using a state space model may be to find a good point estimate of the state variable conditional on observations. It would clearly by desirable for a filter to be able to estimate the whole of the conditional distribution well. For likelihood based (Bayes or frequentist) inference, it is critical to be able to calculate the likelihood of the observations well. The sequence of conditional log likelihoods is $\{\lambda_t, t = 1, \ldots, T\}$, where

$$\lambda_t = \log f(y_t|y_{1,t-1}).$$

For a filtering algorithm to approximate λ_t well, the filter essentially has to capture the whole distribution of x_t given $y_{1,t-1}$. Suppose a filtering algorithm produces an estimate $\hat{\lambda}_t$, then we propose evaluating the algorithm by the *accuracy*, A, defined by

$$A^2 = E[\frac{1}{T}\sum_{t=1}^{T}(\hat{\lambda}_t - \lambda_t)^2].$$

For a stationary model, or at least if $\hat{\lambda}_t - \lambda_t$ is stationary, we have

$$A^2 = E[(\hat{\lambda}_t - \lambda_t)^2].$$

This can be broken down by a bias/variance decomposition, as

$$A^2 = b^2 + V$$

where $b = E[\hat{\lambda}_t - \lambda_t]$ and $V = \mathrm{Var}(\hat{\lambda}_t - \lambda_t)$. For a Monte Carlo filter, let

$$\bar{\lambda}_t = E_{MC}[\hat{\lambda}_t],$$

where E_{MC} is the Monte Carlo expectation. Then we define

$$A^2 = E[E_{MC}[(\hat{\lambda}_t - \lambda_t)^2]].$$

The bias/variance decomposition is now

$$A^2 = b^2 + V + M$$

where $V = \mathrm{Var}(\bar{\lambda}_t - \lambda_t)$ is the *filter variance*, and $M = \mathrm{Var}_{MC}(\hat{\lambda}_t - \bar{\lambda}_t)$ is the *Monte Carlo variance*.

4.1 Advantages and Disadvantages of the Accuracy Score

Any evaluation criterion has plus and minus points. Accuracy, since it concerns ability to approximate the log likelihood, is particularly relevant to inference. It also provides a natural one dimensional measure, even for a vector valued process. Although an inaccurate filter might be able to estimate the state process well, or a parameter well, an accurate filter must capture the whole distribution of the unobserved state process well. The accuracy of a filter may be hard to evaluate, because λ_t is not known unless we already have a perfect filter, with accuracy 0. For the particular nonlinear models of Section 3, λ_t is known. In general, λ_t will be estimated from a filter found to be accurate on such standard problems.

For parameter estimation, a filter has to estimate the likelihood not only under the true model but also in a neighborhood around it. It seems reasonable to suppose that accuracy for a known model is indicative of accuracy for similar models.

4.2 Estimating the Accuracy

Suppose K realizations of a Monte Carlo filter are available, for a stationary state space model, giving rise to estimated likelihood processes $\{\hat{\lambda}_{tk} : t = 1, \ldots, T, k = 1, \ldots, K\}$. Then we can write

$$\hat{\lambda}_{tk} = \lambda_t + b + \epsilon_t + \eta_{tk},$$

where b is the bias, $E[\epsilon_t] = E[\eta_{tk}] = 0$, and ϵ_t is independent of η_{tk}. The filter variance is $V = \mathrm{Var}(\epsilon_t)$ and the Monte Carlo variance is $M = \mathrm{Var}(\eta_{tk})$. A random effects model (Venables & Ripley 2002) could be used to estimate b, M and V, though the standard errors will be too low unless one allows for serial correlation over time. Here we use simpler unbiased estimators,

$$\begin{aligned}
\hat{M} &= \frac{1}{T(K-1)} \sum_{t,k} (\hat{\lambda}_{tk} - \overline{\lambda}_{t\bullet})^2 \\
\hat{b} &= \bar{\lambda}_{\bullet\bullet} \\
\hat{V} &= \frac{1}{TK-1} \sum_{t,k} (\hat{\lambda}_{tk} - \overline{\lambda}_{\bullet\bullet})^2 - \hat{M}
\end{aligned}$$

where $\bar{\lambda}_{t\bullet} = (1/K) \sum_k \hat{\lambda}_{tk}$ and $\bar{\lambda}_{\bullet\bullet} = (1/T) \sum_t \hat{\lambda}_{t\bullet}$. These estimates are repeated for many independent realizations of the state space model, to get improved estimates with standard errors.

5 When There is No Measurement Error

A special case of (1) and (2), with $\tau = 0$, is

$$dx_t = \mu(x_t)dt + \sigma(x_t)dW_t \tag{19}$$
$$y_t = Cx_t, t = 1, 2, \ldots T \tag{20}$$

this case requires special attention, because the law of x_t given $y_{1,t}$ will not have a density with respect to Lebesgue measure on $\mathbb{R}^m$, but instead on the linear space $\{x : Cx = y_t\}$ (recalling that $x_t \in \mathbb{R}^m$, $y_t \in \mathbb{R}^n$, and supposing that $\sigma\sigma'$ is invertible). The $\tau = 0$ case is interesting in its own right, particularly for the discretely observed diffusion situation with $C = I$, the identity matrix. Also, an algorithm that can handle $\tau = 0$ may be expected to deal with small values of τ. The standard particle filter (PF), described in Section 1.2, does not work for $\tau = 0$ since $f(y_t|x_t = X_t^P) = 0$ with probability 1. The extended Kalman filter (EKF) and particle Kalman filter (PKF) function as usual. The conditional particle filter (CPF) also works, but some extra care is required. Let P be the law of $\{x_t|x_0\}$, for $t \in [0, 1]$, and $\hat{P}$ the law of $\{x_t|x_0, Cx_1 = y_1\}$, for $t \in [0, 1]$. Then $dP/d\hat{P}$ cannot be evaluated using Girsanov's theorem (Oksendal 1998, Theorem 8.6.5) since Novikov's condition fails to apply. A way around this problem is to define P_δ and $\hat{P}_\delta$ as the laws of x_t and $\hat{x}_t$ restricted to $t \in [0, 1 - \delta]$. Now $dP_\delta/d\hat{P}_\delta$ does exist, and we can use

$$f(y_1|x_0) = E_{\hat{P}_\delta}\left[f(y_1|x_{1-\delta}{=}\hat{x}_{1-\delta})\frac{dP_\delta}{d\hat{P}_\delta}\{\hat{x}_t\}\right]. \tag{21}$$

Fortunately, in CPF this complication does not much effect the Monte Carlo part of the problem, which is to simulate from a law $\tilde{P}$ close to $\hat{P}$ and estimate $f(y_1|x_0)$ using

$$f(y_1|x_0) = E_{\tilde{P}_\delta}[f(y_1|x_{1-\delta}{=}\tilde{x}_{1-\delta})\frac{dP_\delta}{d\tilde{P}_\delta}]. \tag{22}$$

In Section 5.1 we see that as long as $\tilde{P}$ is similar to $\hat{P}$, the two terms in the expectation of (22) are of order $(1/\sqrt{\delta}) \times \sqrt{\delta}$ for typical sample paths. A small value of δ contributes numerical instability while adding relatively little Monte Carlo variation, since the numerical instabilities cancel. Sections 5.1, 5.2 and 5.3 investigate the choice of δ.

5.1 Example (Brownian bridge)

Consider the system

$$dx_t = dW_t$$

$$x_0 = 0, \quad y_1 = x_1 = 0,$$

which is a case of (19) and (20). Then,

$$\frac{dP_\delta}{d\hat{P}_\delta}(\xi) = \exp\left\{2\int_0^{1-\delta} \frac{\xi_t d\xi_t}{1-t} + \int_0^{1-\delta} \frac{\xi_t^2}{(1-t)^2} dt\right\} \tag{23}$$

As δ approaches 0, there are numerical difficulties evaluating the right hand side of (21), since $f(y_1 \mid x_{1-\delta}$ is approximately the density of a $N(x_{1-\delta}, \delta)$ variable. Thus for typical sample paths under $\hat{P}_\delta$, $f(y_1|x_{1-\delta}{=}\tilde{x}_{1-\delta})$ is of order $1/\sqrt{\delta}$. Novikov's condition, in this example, concerns variables

$$\nu_\delta(\hat{x}) = \frac{1}{2}\int_0^{1-\delta} \frac{\hat{x}_t^2}{(1-t)^2} dt$$

The condition fails because

$$E_{\hat{P}}[e^{\nu_0}] = \infty$$

Conditional on ν_δ, in this example, $dP_\delta/d\tilde{P}_\delta$ has distribution

$$\exp\{\sqrt{2\nu_\delta}Z - \nu_\delta\}$$

where $Z \sim N(0,1)$, so keeping track of how ν_δ increases as δ approaches 0 may help to choose a reasonable value of δ.

5.2 Example (Brownian Bridge Continued)

A test problem is to calculate $f(y_1|x_0)$ for (5.1) and (5.1), by using (22). The correct answer is $1/\sqrt{2\pi} = 0.3989$. $\tilde{x}_t$ was taken to be an Euler approximation to

$$d\hat{x}_t = \frac{-\hat{x}_t}{1-t}dt + dW_t,$$

based on discretizing $[0,1]$ into N_t equal intervals,and taking $\delta = 1/N_t$.

Table 2 shows that the SD of the estimate of $f(y_1|x_0)$ increases remarkably slowly as δ becomes small. The importance weights, $dP_\delta/d\hat{P}_\delta$ have a very long tailed distribution for small δ, but the large values of $dP_\delta/d\hat{P}_\delta$ match small values of $f(y_1|x_{1-\delta} = \hat{x}_{1-\delta})$. Since the observed mean of $dP_\delta/d\hat{P}_\delta$ is less than one for small δ, the extreme tail of the distribution is not even being sampled here, but that is not affecting the mean of the estimates $\hat{f}(y_1|x_0)$ of $f(y_1|x_0)$.

	N_t	$\hat{f}(y_1\|x_0)$	$dP_\delta/d\hat{P}_\delta$	$f(y_1\|\hat{x}_{1-\delta})$	ν_δ
mean	5	0.399	1.00	0.571	0.324
SD	5	0.216	2.63	0.279	0.371
Mean	10	0.399	0.92	0.787	0.640
SD	10	0.235	3.68	0.404	0.673
Mean	25	0.396	0.83	1.238	1.091
SD	25	0.248	8.60	0.640	1.045
Mean	50	0.398	0.68	1.745	1.408
SD	50	0.251	18.0	0.913	1.255

Table 2: Monte Carlo estimates, $\hat{f}(y_1|x_0)$, of $f(y_1|x_0)$ using CPF for a Brownian bridge. Based on a simulation study with sample size 10^4. Note that $\delta = 1/N_t$ here.

Note that in this example $\hat{f}(y_1|x_0)$ is unbiased, since the numerical solution of the SDE for x_t is exact – the Euler method for Brownian motion gives correct finite dimensional distributions. This reminds us that the bias is due to the error in the numerical solution for x_t, not in the approximation of $\hat{x}_t$ by $\tilde{x}_t$. For the Brownian bridge, $\hat{x}_t$, the Euler solution is not exact.

5.3 Example (A Nonlinear Diffusion)

For a slightly more challenging example, we consider the nonlinear diffusion (14) observed discretely with no measurement error. Some results are presented in Table 3. The filter error, $\sqrt{V}$, decreases as N_t increases and as $\delta = 1 - N_s/N_t$ decreases. The Monte Carlo error, $\sqrt{M}$, increases as δ decreases, but decreases proportional to $(N_r)^{-1/2}$ as the number of Monte Carlo replications, N_r, increases.

6 Discussion

There are several potential refinements and extensions of the conditional particle methods developed in this paper, which should be investigated if/when these methods are found to make a useful addition to current filtering techniques.

- If $f(x_t|y_{1,t})$ and/or $f(y_{t+1}|x_t)$ are sufficiently smooth, as functions of x_t, it may be possible to reduce Monte Carlo variation by applying some smoothing technique. Stavropoulos & Titterington (2001) investigated this idea for uncondi-

	N_t	N_s	$\sqrt{V} \times 10^2$	$\sqrt{M} \times 10^2$	$A \times 10^2$
EKF			61		62
CPF	5	4	13.8	4.51	14.6
CPF	10	9	7.62	5.05	9.21
CPF	10	8	10.20	3.76	10.93
CPF	20	19	4.12	5.48	6.90
CPF	20	18	5.07	4.23	6.63
CPF	20	16	8.63	3.33	9.32

Table 3: Comparing EKF with CPF for 14, observed discretely with no measurement error. The bias, b, was always small and negative, and is not reported. For CPF, $N_r = 200$. Setting $N_r = 800$ results in indistinguishable b and $\sqrt{V}$, and halves $\sqrt{M}$ (not reported). Note that $\delta = 1 - N_s/N_t$. The computational effort is proportional to $N_t \times N_r$. CPF provides a marked improvement on EKF (which here is simply a local linearization of the diffusion process), even for modest values of N_t and N_r.

tional particle filters. These modifications must necessarily require more careful implementation and the selection of smoothing parameters.

- Section 2.3 suggested that more work could be done to develop strategies for weighting and/or resampling.

- Considerable computational gains may be available for partially linear/Gaussian systems. This occurs where the state process can be split into two components, $x_t = (x_t^{(1)}, x_t^{(2)})$, and $x_t^{(1)}$ given x_{t-1} and $x_t^{(2)}$ is conditionally linear and Gaussian. Investigations of unconditional particle filtering in this situation include (Liu, Chen & Logvinenko 2001) and (Murphy & Russell 2001).

- A goal of developing more efficient filtering algorithms is to apply them to increasingly complex models, with large state spaces. One example is the so-called data assimilation problem, where observations are used to adjust a computer simulation of a large dynamical system. Data assimilation problems, such as combining geophysical models with observatio stations and sattelites for weather forecasting, may be written as state space models. Methods similar to particle filters have been developed for weather prediction (Evensen & van Leeuwen 1996), but the high dimension of the state and observation spaces requires modification of the algorithm.

Although there is scope for refinement, this paper provides readily applicable algorithms for filtering, smoothing, prediction and likelihood evaluation for a major class of state space models. The motivation given for the algorithms, and the test problems solved, suggest that these algorithms may be more effective than available alternatives.

A Local Linearization - Zeroth Order

Suppose that $\{x_t\}$ is a Brownian motion with drift in $\mathbb{R}^n$,

$$dx_t = \mu dt + \sigma dW_t.$$

We wish to find the diffusion $\{\tilde{x}_t\}$ arising from conditioning $\{x_t\}$ on $y = Cx_T + D + \eta$, $\eta \sim N(0, \tau\tau')$, for $t \leq T$. Write

$$d\tilde{x}_t = \tilde{\mu}(\tilde{x}_t, t)dt + \sigma dW_t,$$

using the knowledge that the conditional process is a diffusion with the same infinitesimal variance as the original diffusion. Let $\epsilon = x_{t+\delta} - E[x_{t+\delta}] \dot{\sim} N(0, \delta\sigma\sigma')$, and note that

$$\mathrm{Cov}(\epsilon, y) = \mathrm{Cov}(\epsilon, C\epsilon) \doteq \delta\sigma\sigma'C',$$

where $\doteq$ and $\dot{\sim}$ indicate linearizations for small δ. Then

$$\begin{aligned} E[\epsilon|y] &\doteq \Sigma_{\epsilon y}\Sigma_{yy}^{-1}(y - E[y]) \\ &\doteq \delta\sigma\sigma'C'(C(T-t)\sigma\sigma'c' + \tau\tau')(y - D - C(x_t + (T-t)\mu)). \end{aligned}$$

Taking $\delta \downarrow 0$ gives

$$\tilde{\mu}(\tilde{x}_t, t) = \mu + \sigma\sigma'C'(C(T-t)\sigma\sigma'c' + \tau\tau')(y - D - C(\tilde{x}_t + (T-t)\mu)).$$

Note that if C and σ are invertible (e.g. if $n = 1$) and $\tau = 0$ then

$$d\tilde{x}_t = \frac{1}{\tau - t}(C^{-1}(y - D) - \tilde{x}_t)dt + \sigma dW_t.$$

In this case, $\tilde{x}_t$ is a Brownian bridge and the drift does not depend on μ.

B Local Linearization - First Order

Suppose that $\{x_t\}$ is an Ornstein-Uhlenbeck process in $\mathbb{R}^n$,

$$dx_t = \alpha(x_t - \mu)dt + \sigma dW_t.$$

We wish to find the diffusion $\{\tilde{x}_t\}$, conditional on $y = Cx_T + \eta$, $\eta \sim N(0, \tau\tau')$, for $t \leq T$. Let $\epsilon = x_{t+\delta} - E[x_{t+\delta}] \dot{\sim} N(0, \delta\sigma\sigma')]$, and note that

$$\mathrm{Cov}(\epsilon, y) \doteq \mathrm{Cov}(\epsilon, Ce^{\alpha(T-t)}\epsilon) \doteq \delta\sigma\sigma' e^{\alpha'(T-t)}C'$$

Then,

$$\begin{aligned} E[x_{t+\delta}|y] - E[x_{t+\delta}] &\doteq \Sigma_{\epsilon y}\Sigma_{yy}^{-1}(y - E[y]) \\ &\doteq \delta\sigma\sigma' e^{\alpha' t}C'(C\Sigma_t C' + \tau\tau')^{-1}(y - C[\mu + e^{\alpha(T-t)}(x_t - \mu)]) \end{aligned}$$

where

$$\begin{aligned} \Sigma_t &= \mathrm{Var}(x_T|x_t) \\ &= E[\int_t^T e^{\alpha(T-s)}\sigma dW_s \{\int_t^T e^{\alpha(T-s)}\sigma dW_s\}'] \\ &= \int_0^{T-t} e^{\alpha s}\sigma\sigma' e^{\alpha' s} ds \\ &= [e^{\alpha s} M e^{\alpha' s}]_0^{T-t} \\ &= e^{\alpha(T-t)} M e^{\alpha'(T-t)} - M. \end{aligned}$$

for M defined by

$$\alpha M + M\alpha' = \sigma\sigma'.$$

Then $\tilde{x}_t$ has SDE

$$d\tilde{x}_t = (\alpha(\tilde{x}_t - \mu) + \nu_t(\tilde{x}_t))dt + \sigma dW_t.$$

where

$$\nu_t(x_t) = \sigma\sigma' C' e^{\alpha t}(C\Sigma_t C' + \tau\tau')^{-1}(y - C[\mu + e^{\alpha(T-t)}(x_t - \mu)]).$$

C The Linearization for Section 3.1

First set

$$\begin{aligned}
\mu_t &= -\alpha h^{-1}(x_t^f)h'(h^{-1}(x_t^f)) + \frac{1}{2}h''(h^{-1}(x_t^f)) \\
\mu_t' &= \frac{d}{dx}[-\alpha h^{-1}(x)h'(h^{-1}(x)) + \frac{1}{2}h''(h^{-1}(x))]|_{x=x_t^f} \\
&= -\alpha - \frac{\alpha h^{-1}(x_t^f)h''(h^{-1}(x_t^f))}{h'(h^{-1}(x_t^f))} + \frac{h'''(h^{-1}(x_t^f))}{2h'(h^{-1}(x_t^f))}
\end{aligned}$$

Then,

$$\begin{aligned}
A_t &= 1 + \mu_t' \\
B_t &= \mu_t - \mu_t' x_t^f \\
\Phi_t^2 &= [h'(h^{-1}(x_t^f))]^2 \\
x_{t+1}^p &= A_t x_t^f + B_t \\
h^{-1}(x_{t+1}) &= h^{-1}(x_{t+1}^p + (x_{t+1} - x_t^p)) \\
&\doteq h^{-1}(x_{t+1}^p) + [1/h'(h^{-1}(x_{t+1}^p))](x_{t+1} - x_{t+1}^p) \\
C_{t+1} &= 1/h'(h^{-1}(x_{t+1}^p)) \\
D_{t+1} &= h^{-1}(x_{t+1}^p) - \frac{x_{t+1}^p}{h'(h^{-1}(x_{t+1}^p))}
\end{aligned}$$

References

Anderson, B. D. & Moore, J. B. (1979), *Optimal Filtering*, Prentice-Hall, New Jersey.

Barndorff-Nielsen, O. E., Kosch, T. M. & Resnick, S. I. (2001), *Lévy process, Theory and applications*, Birkhäuser, New York.

Bickel, P. J. & Ritov, Y. (1996), 'Inference in hidden Markov models, I: Local asymptotic normality in the stationary case', *Bernoulli* **2**, 199–228.

Bickel, P. J., Ritov, Y. & Ryden, T. (1998), 'Asymptotic normality of the maximum-likelihood estimator for general hidden Markov models', *Ann. Statist.* **26**, 1614–1635.

Brown, E. N., Frank, L. M., Tang, D., Quirk, M. C. & Wilson, M. A. (1998), 'A statistical paradigm for neural spike train decoding applied to position prediction from ensemble firing patterns of rat hippocampal place cells', *J. Neuroscience* **18**, 7411–7425.

Crisan, D. & Lyons, T. (1997), 'Nonlinear filtering and measure-valued processes', *Probability Theory and Related Fields* **109**(2), 217–244.

Del Moral, P. (1996), 'Nonlinear filtering using random particles', *Theory Probab. Appl.* **40**, 690–701.

Doucet, A., de Freitas, N. & Gordon, N. J. (2001), *Sequential Monte Carlo methods in practice*, Springer, New York.

Durbin, J. & Koopman, S. J. (2001), *Time series analysis by state space methods*, Oxford University Press.

Elerian, O., Chib, S. & Shephard, N. (2001), 'Likelihood inference for discretely observed nonlinear diffusions', *Econometrica* **69**, 959–993.

Evensen, G. & van Leeuwen, P. J. (1996), 'Assimilation of geostat altimeter data for the agulhas current using the ensemble kalman filter', *Monthly Weather Review* **124**, 58–96.

Gordon, N., Salmond, D. J. & Smith, A. F. M. (1993), 'Novel approach to nonlinear/non-gaussian bayesian state estimation', *IEE Proceedings–F* **140**(2), 107–113.

Hammersley, J. M. & Morton, K. W. (1954), 'Poor man's Monte Carlo', *J. R. Statist. Soc. B* **16**, 23–38.

Harvey, A. C. (1989), *Forecasting, structural time series models and the Kalman filter*, Cambridge University Press.

Hürzeler, M. & Künsch, H. R. (2001), Approximating and maximising the likelihood for a general state-space model, *in* A. Doucet, N. de Freitas & N. J. Gordon, eds, 'Sequential Monte Carlo methods in practice', Springer, New York, pp. 159–175.

Ionides, E. L., Fang, K. S., Isseroff, R. R. & Oster, G. F. (2003), 'Stochastic models for cell motion and taxis', *J. Math. Biol.* p. In press.

Isard, M. & Blake, A. (1996), Contour tracking by stochastic propagation of conditional density, *in* 'European Conference on Computer Vision', Cambridge, UK, pp. 343–356.

Jensen, J. L. & Petersen, N. V. (1999), 'Asymptotic normality of the maximum likelihood estimator in state space models', *Ann. Statist.* **27**, 514–535.

Kalman, R. E. (1960), 'A new approach to linear filtering and prediction problems', *J. Basic Eng.* **82**, 35–45.

Kitagawa, G. (1996), 'Monte Carlo filter and smoother for non-Gaussian nonlinear state space models', *J. Comput. Graph. Statist.* **5**, 1–25.

Kitagawa, G. & Sato, S. (2001), Monte carlo smoothing and self-organising state-space model, *in* A. Doucet, N. de Freitas & N. J. Gordon, eds, 'Sequential Monte Carlo methods in practice', Springer, New York, pp. 177–195.

Kloeden, P. E. & Platen, E. (1999), *Numerical Soluion of Stochastic Differential Equations*, 3rd edn, Springer, New York.

Lander, E. S. & Green, P. (1987), 'Construction of multilocus genetic linkage maps in humans', *Proc. Natl. Acad. Sci. USA* **84**, 2363–2367.

Liu, J. S., Chen, R. & Logvinenko, T. (2001), A theoretical framework for sequential importance sampling with resampling, *in* A. Doucet, N. de Freitas & N. J. Gordon, eds, 'Sequential Monte Carlo methods in practice', Springer, New York, pp. 225–242.

Murphy, K. & Russell, S. (2001), Rao-blackwellised particle filtering for dynamic bayesian networks, *in* A. Doucet, N. de Freitas & N. J. Gordon, eds, 'Sequential Monte Carlo methods in practice', Springer, New York, pp. 499–515.

Oksendal, B. (1998), *Stochastic Differential Equations*, 5th edn, Springer, New York.

Ozaki, T. (1992), 'A bridge between nonlinear time series models and nonlinear stochastic dynamic systems: a local linearization approach', *Statist. Sinica* **2**, 113–135.

Pitt, M. K. & Shepard, N. (2001), Auxillary variable based particle filters, *in* A. Doucet, N. de Freitas & N. J. Gordon, eds, 'Sequential Monte Carlo methods in practice', Springer, New York, pp. 273–293.

Rabiner, L. R. (1989), A tutorial on hidden Markov models and selected applications in speech recognition, *in* 'Proc. IEEE', Vol. 77, pp. 257–285.

Roberts, G. O. & Stramer, O. (2001), 'On inference for partially observed nonlinear diffusion models using the Metropolis-Hastings algorithm', *Biometrika* **88**, 603–621.

Rosenbluth, M. N. & Rosenbluth, A. W. (1955), 'Monte Carlo calculation of the average extension of molecular chains', *J. Chem. Physics* **23**, 356–359.

Rubin, D. B. (1987), 'A noniterative sampling/importance resampling alternative to the data augmentation algorithm for creating a few imputations when fractions of missing information are modest: the sir algorithm. comment on the calculation of posterior distributions by data augmentation, by M. A. Tanner and W. H. Wong', *J. Amer. Statist. Assoc.* **82**, 543–546.

Shephard, N. & Pitt, M. K. (1997), 'Likelihood analysis of non-Gaussian measurement time series', *Biometrika* **84**, 653–667.

Shumway, R. H. & Stoffer, D. S. (2000), *Time Series Analysis and Its Applications*, Springer, New York.

Stavropoulos, P. & Titterington, D. M. (2001), Improved particle filters and smoothing, *in* A. Doucet, N. de Freitas & N. J. Gordon, eds, 'Sequential Monte Carlo methods in practice', Springer, New York, pp. 295–317.

Venables, W. N. & Ripley, B. D. (2002), *Modern Applied Statistics with S-Plus*, 4th edn, Springer, New York.