Smoothing Algorithms for Variable Rate Models

Pete Bunch, Simon Godsill, *Member, IEEE,*

***Abstract*—The abstract goes here.**

***Index Terms*—**

I. Introduction

THE objective of sequential Bayesian inference is to estimate an imperfectly observed quantity as it varies over time. This is accomplished through the use of probabilistic models for the state evolution and measurement processes. Often, the latent state is a continuously varying quantity, whereas the observations are made at a discrete set of times. In these circumstances, it is simplest to discretise the state onto the same set of times as the observations. When the system is also Markovian, this leads to the standard "fixed rate" hidden Markov model (HMM). The standard HMM is poorly suited to systems where the state evolution contains discontinuities; for example, the price of a financial asset which may display large jumps at random times between periods of diffusion-like behaviour, or the kinematic state of a manoeuvring vehicle which may have sudden changes in the acceleration when turns begin or end. Such problems can be handled more naturally using a "variable rate" model, in which the state dynamics are conditioned upon a set of random changepoints which characterise transitions in behaviour.

In a variable rate model, the set of changepoints and associated parameters are modelled as a marked point process (MPP), the mathematical properties of which are thoroughly set out in [1]. Conditional upon this MPP the state evolves according to some benign dynamics. In [2], [3], the conditional state evolution is treated as deterministic, while in [4], [5] a conditionally linear-Gaussian state model is considered.

The posterior distribution for the changepoint MPP is inherently nonlinear, and cannot be calculated analytically. Instead, inference must be conducted using numerical approximations. The particle filter (introduced by [6]) and smoother (see [7], [8]) are schemes which approximate a posterior distribution using a set of samples, or "particles", drawn sequentially from it. A thorough introduction to particle filtering and smoothing methods can be found in [9], [10]. In [2]–[4], the particle filter was adapted for use with variable rate models, resulting in the variable rate particle filter (VRPF).

The VRPF allows the changepoint sequence – and hence the current state – to be estimated sequentially as observations are received. However, estimates can often be improved later once further observations have been made. In this paper, we address the problem of smoothing in variable models, i.e. the estimation of the changepoint sequence and latent state given all the observations. This is achieved with an efficient backward sweep through the observations, in a similar manner to the method for standard HMMs described in [8]. Two new schemes are introduced: one for conditionally linear-Gaussian models which exploits the method of Rao-Blackwellisation; and a second for use with conditionally-deterministic models which uses an augmented target distribution in the style of an SMC sampler [11].

We introduce the general structure of variable rate models in section II and discuss two particular types, those with conditionally linear-Gaussian and conditionally deterministic dynamics. The VRPF is reviewed in section III, and new variable rate smoothing algorithms are described in section IV. In section **??**, particular examples of variable rate models are presented and their performance demonstrated in a series of simulations.

P. Bunch and S. Godsill are with the Department of Engineering, Cambridge University, UK. email: {pb404,sjg30}@cam.ac.uk

Manuscript received January 01, 1901; revised January 02, 1901.

II. Variable Rate Models

We consider a general model from time 0 to T, between which observations, $\{y_1 \dots y_N\}$, are made at times $\{t_1 \dots t_N = T\}$. During this period, an unknown number of changepoints, K, occur at times $\{\tau_0 = 0, \tau_1 \dots \tau_K\}$, each with associated changepoint parameters, $\{u_0, u_1 \dots u_K\}$. The pairs $\{\tau_k, u_k\}$ are the elements of an marked point process (MPP). We will refer to the times of the MPP as the "change-time sequence" and marks as "parameter sequence", and both together as the "changepoint sequence". The latent state is a continuous-time process denoted $x(t)$. Discrete sets containing multiple values over time will be written as, e.g. $y_{1:n} = \{y_1 \dots y_n\}$.

The objective for inference will be to estimate the changepoint sequence. This will be denoted as $\theta = \{\tau_{0:K}, u_{0:K}\}$. At a particular time t_n, the sequence will be divided into past $\theta_n = \{\tau_j, u_j \forall j : 0 \leq \tau_j < t_n\}$, and future $\theta_n^+ = \{\tau_j, u_j \forall j : t_n \leq \tau_j < T\}$. It will also be useful to define a variable for the changepoints which occur in the interval $[t_{n-1}, t_n)$, $\theta_{n \setminus n-1} = \{\tau_j, u_j \forall j : t_{n-1} \leq \tau_j < t_n\}$.

For notational simplicity, the following counting variables are introduced to keep track of the most recent changepoint to have occurred,

$$K(t) = \max(k : \tau_k < t) \tag{1}$$

$$K_n = K(t_n). \tag{2}$$

The changepoint sequence is assumed to be a Markov process.

$$\{\tau_k, u_k\} \sim p(u_k|\tau_k, \tau_{k-1}, u_{k-1}) p(\tau_k|\tau_{k-1}, u_{k-1}) \tag{3}$$

The changepoint density will be constructed such that $P(\tau_k < \tau_{k-1}) = 0$.

In the manner of [3], a survivor function is defined as the probability that no new changepoint occurs before a given time,

$$S(\tau_k, u_k, t) = P(\tau_{k+1} > t | \tau_k, u_k) = 1 - \int_{\tau_k}^{t} p(\xi|\tau_k, u_k) d\xi. \quad (4)$$

It is now possible to write down a prior for the changepoint sequence, where we use the convention that $\tau_0 = 0$. The existence of such a density for a MPP is addressed in [1].

$$p(\theta_n) = S(\tau_{K_n}, t_n) p(u_0) \prod_{k=1}^{K_n} p(\tau_k, u_k | \tau_{k-1}, u_{k-1}) \quad (5)$$

A. Conditionally Linear-Gaussian Models

The first class of variable rate models to be considered is those whose state dynamics are linear-Gaussian conditional on the changepoint sequence. Such a model may be discretised onto the set of observation times in exactly the same manner as a standard the standard HMM.

$$x_n = A_n(\theta_n) x_{n-1} + w_n \quad (6)$$

$$y_n = C_n(\theta_n) x_n + v_n, \quad (7)$$

where,

$$w_n \sim \mathcal{N}(w_n | 0, Q_n(\theta_n)) \quad (8)$$

$$v_n \sim \mathcal{N}(v_n | 0, R_n(\theta_n)). \quad (9)$$

In addition, the prior state distribution should be Gaussian, with known mean and variance.

$$x_0 \sim \mathcal{N}(x_0 | m_0, P_0). \quad (10)$$

If the changepoint sequence is known, or has been estimated, then the state values, x_n may be inferred using optimal Kalman filtering and smoothing recursions. As well as the basic Kalman filter [12], the Rauch-Tung-Striebel (RTS) smoother [13] and two-filter smoother [14] will prove useful for this step.

Conditionally linear-Gaussian variable rate models were introduced in [4] for a financial inference algorithm. Changepoints correspond to jumps in the value or trend of a security, at which points the process covariance, $Q_n(\theta_n)$, is inflated.

B. Conditionally Deterministic Models

The second class of variable rate models for consideration is those in which the state is completely specified by the changepoint sequence, with no additional random components. Such a process is commonly referred to as "piecewise-deterministic", as the latent state follows a deterministic path between changepoints. In this case, it is not necessary to discretise the state – it may be kept as a continuous variable.

In general, the state dynamics will be governed by a differential equation which may depend on the entire changepoint sequence. Here we assume that only the most recent changepoint is significant.

$$dx(t) = \mathrm{f}(x(t), \tau_{K(t)}, u_{K(t)}). \quad (11)$$

By introducing a new sequence, $\{x_0, x_1 \dots x_K\}$, which denotes the value of the state at each changepoint (i.e. $x(\tau_k)$), and assuming that an analytic solution exists, a state transition function may be found,

$$x(t) = f(x_{K_n}, v_{K_n}, \tau_{K_n}, t), \tau_{K_n} < t <= \tau_{K_n+1}. \quad (12)$$

By choosing $t = \tau_{K_n+1}$, this equation specifies the state at the next changepoint time. Similarly, by choosing $t = t_n$, the state at the observation times may be calculated. These points will be denoted $\hat{x}_n$. To complete the model, a probabilistic measurement model must be chosen for the observation process, $p(y_n|\hat{x}_n)$.

For convenience, we assume that x_0 is known in the following sections. This means that $x(t)$ may be calculated deterministically for all t given θ. However, this condition is easily relaxed by including x_0 as a random variable in the posterior distribution.

Target tracking algorithms are commonly based upon fixed rate models (see, e.g. [15] for a thorough survey), in which the target kinematics (position, velocity, etc.) are estimated at a set of fixed times at which observations (e.g. radar measurements) are made. In [2], [4], variable rate models were introduced for tracking, in which the state trajectory is divided up by a set of changepoints between which the motion follows a deterministic path governed by motion parameters (accelerations, etc.) which are fixed for that division.

III. The Variable Rate Particle Filter

The variable rate particle filter (VRPF) is described in [2], [4]. The objective of the algorithm is to sequentially estimate the posterior distribution of the changepoint sequence, $p(\theta_n|y_{1:n})$, at each time t_n. This distribution may be expanded using Bayes' rule.

$$p(\theta_n|y_{1:n}) \propto p(y_n|\theta_n, y_{1:n-1}) p(\theta_{n\setminus n-1}|\theta_{n-1}) p(\theta_{n-1}|y_{1:n-1}) \quad (13)$$

The transition term, $p(\theta_{n\setminus n-1}|\theta_{n-1})$, can be constructed in a similar manner to (5) [1], as a product of density terms for each new changepoint and a survivor function term. In the particular (but common) case that no new changepoints occur within the interval, the density consists of a probability mass on the empty set, with weight $P(\tau_{K_n+1} > t_n | \tau_{K_n+1} > t_{n-1}, \tau_{K_n})$.

$$p(\theta_{n\setminus n-1}|\theta_{n-1}) = \begin{cases} S(\tau_{K_n}, t_n) \prod_{j: t_{n-1} \le \tau_j < t_n} p(\tau_j, u_j | \tau_{j-1}, u_{j-1}, \tau_j > t_{n-1}) & K_n > K_{n-} \\ S(\tau_{K_n}, t_n) / S(\tau_{K_n}, t_{n-1}) & K_n = K_{n-} \end{cases}$$

For all but the first changepoint in the interval, the density is given by the prior model of (3). For the first changepoint, indexed by $k = K_{n-1} + 1$, we must account for the fact that a changepoint cannot occur before t_{n-1},

$$p(\tau_k, u_k | \tau_{k-1}, u_{k-1}, \tau_k > t_{n-1}) = \frac{1}{S(\tau_{k-1}, t_{n-1})} \begin{cases} p(\tau_k, u_k | \tau_{k-1}, u_{k-1}) & \tau_k > t_{n-1} \\ 0 & \tau_k < t_{n-1} \end{cases} \quad (15)$$

Practically, because changepoints will be relatively rare events, it is not likely that more than one new changepoint will occur between t_{n-1} and t_n.

The target distribution of (13) cannot be calculated analytically, but may be approximated numerically. A particle filter is an algorithm for approximating a probability distribution using a set of weighted samples (or "particles") drawn from that distribution. In this case, each particle will be a set of changepoint times and parameters.

$$\hat{p}(\theta_n|y_{1:n}) = \sum_j w_n^{(j)} \delta_{\theta_n^{(j)}}(\theta_n) \tag{16}$$

where $\delta_x(X)$ is a dirac probability mass at $X = x$.

The particle filter works recursively. At the nth step, a particle, $\theta_{n-1}^{(i)}$, is first resampled from those approximating the filtering distribution at the $(n-1)$th step, using an appropriately chosen set of proposal weights.

$$q(\theta_{n-1}) = \sum_j v_{n-1}^{(j)} \delta_{\theta_{n-1}^{(j)}}(\theta_{n-1}) \tag{17}$$

The choice of weights determines the type of resampling used. The simplest choice, $v_{n-1}^{(j)} = 1/N_F$ (where N_F is the number of filter particles) may be achieved by simply omitting this step all together and using the particles of $\hat{p}(\theta_{n-1}|y_{1:n-1})$. This, however, leads to degeneracy of the particle weights over time. Conventional resampling is achieved by using $v_{n-1}^{(j)} = w_{n-1}^{(j)}$. Any other choice results in an auxiliary particle filter [16]. For further discussion of resampling, see [9], [10], [17].

Next, an extension to the changepoint sequence, $\theta_{n\backslash n-1}^{(i)}$, is proposed from an importance distribution, $q(\theta_{n\backslash n-1}|\theta_{n-1}^{(i)}, y_n)$, and concatenated to θ_{n-1} to create an estimate of θ_n. Finally, the particle is weighted according to the ratio of the target and proposal densities.

$$\begin{aligned} w_n^{(i)} &= \frac{p(\theta_n^{(i)}|y_{1:n})}{q(\theta_n)} \\ &\propto \frac{p(y_n|\theta_n, y_{1:n-1})p(\theta_{n\backslash n-1}|\theta_{n-1})p(\theta_{n-1}|y_{1:n-1})}{q(\theta_{n-1})q(\theta_{n\backslash n-1}|\theta_{n-1}^{(i)}, y_n)} \\ &= \frac{w_{n-1}^{(i)}}{v_{n-1}^{(i)}} \times \frac{p(y_n|\theta_n, y_{1:n-1})p(\theta_{n\backslash n-1}|\theta_{n-1})}{q(\theta_{n\backslash n-1}|\theta_{n-1}^{(i)}, y_n)} \end{aligned} \tag{18}$$

The normalisation may be enforced by scaling the weights so that they sum to 1.

For the most basic "bootstrap" [6] form of the VRPF, $\theta_{n\backslash n-1}$ may be proposed from the prior transition density (14). This can be achieved by sampling new changepoints sequentially from the transition model (3) (apart from the first which is sampled from (15)) until one falls after the current time, t_n. This final future changepoint is discarded. (This process can be thought of as sampling the entire future changepoint sequence from t_{n-1} onwards, and then marginalising those which fall after t_n.) The bootstrap proposal leads to the usual simplification of the weight formula.

$$w_n^{(i)} = \frac{w_{n-1}^{(i)}}{v_{n-1}^{(i)}} \times p(y_n|\hat{x}_n) \tag{19}$$

The choice of proposal weights, $\{v_{n-1}^{(i)}\}$, requires particular attention in the design of VRPFs. In some models a changepoint may not have an immediate effect on the observations, especially if a jump occurs in some quantity which is only observed via its integral, e.g. if there is a jump in the acceleration of a moving object, yet only the position is measured, the change will not be apparent until several more observations have been made. In the meantime, particles which contain a changepoint at the correct time may all have been removed by the resampling process. To avoid this loss of good particles, proposal weights should be chosen which preserve a significant number of low-weight particles. One scheme which has been found to work well is described in [2], in which proposal weights are given by:

$$v_{n-1}^{(i)} \propto \max(1, N_F w_{n-1}^{(i)}) \tag{20}$$

It only remains to consider the likelihood term required for evaluation of the importance weights, $p(y_n|\theta_n, y_{1:n-1})$. The form of this term depends on the particular model under consideration. In the following sections, the likelihood expressions for the conditionally linear-Gaussian and deterministic cases are considered.

A. Conditionally Linear-Gaussian Likelihoods

For a conditionally linear-Gaussian model, the required likelihood term, $p(y_n|\theta_n, y_{1:n-1})$ is the predictive distribution estimated by the Kalman filter. Conveniently, the Kalman filter also provides us with an estimate of the current state given the changepoint sequence and all the preceding observations. It follows from the the Gaussian dynamics and prior that these distributions are all Gaussian as well, [18],

$$p(x_n|\theta_n, y_{1:n-1}) = \mathcal{N}(x_n|m_n^-, P_n^-) \tag{21}$$
$$p(x_n|\theta_n, y_{1:n}) = \mathcal{N}(x_n|m_n, P_n) \tag{22}$$
$$p(y_n|\theta_n, y_{1:n-1}) = \mathcal{N}(y_n|\mu_n, S_n), \tag{23}$$

with means and variances given by the following standard recursions (dependence on θ_n suppressed for clarity),

$$m_n^- = A_n m_{n-1} \tag{24}$$
$$P_n^- = A_n P_{n-1} A_n^T + Q_n \tag{25}$$
$$\mu_n = C_n m_n^- \tag{26}$$
$$S_n = C_n P_n^- C_n^T + R_n \tag{27}$$
$$K_n = P_n^- C_n^T S_n^{-1} \tag{28}$$
$$m_n = m_n^- + K_n(y_n - \mu_n) \tag{29}$$
$$P_n = P_n^- - K_n S_n K_n^T. \tag{30}$$

This completes the requirements for the particle filter, resulting in the final algorithm shown in Fig. 1. The concept of estimating the posterior for just the changepoint sequence with a particle filter rather than the joint posterior over changepoints and linear states is an example of Rao-Blackwellisation (see, e.g. [19], [20]), so this algorithm is known as the Rao-Blackwellised variable rate particle filter (RB-VRPF) [4], [5].

For each i, initialise particle changepoint sequence with $\{\theta_0^{(i)}\} \leftarrow \{\tau_0^{(i)}, u_0^{(i)}$, where $\tau_0^{(i)} = 0$ and $\{u_0^{(i)}\} \sim p(u_0)$.
For each i, initialise particle sufficient statistics, $m_0^{(i)}$ and $P_0^{(i)}$ with prior values.
for $n = 1 \dots N$ **do**
  for $i = 1 \dots N_F$ **do**
    Sample a changepoint history $\theta_{n-1}^{(i)} \sim \sum_j v_{n-1}^{(j)} \delta_{\theta_{n-1}^{(j)}}(\theta_{n-1})$.
    Propose a sequence extension $\theta_{n\backslash n-1}^{(i)} \sim q(\theta_{n\backslash n-1}|\theta_{n-1}^{(i)})$.
    Add extension to changepoint sequence $\theta_n^{(i)} \leftarrow \theta_{n-1}^{(i)} \cup \theta_{n\backslash n-1}^{(i)}$.
    Predict state mean and covariance $\mu_n^{(i)}$ and $S_n^{(i)}$ using (24) to (27).
    Calculate weight $w_n^{(i)}$ using (18).
    Update state mean and covariance $m_n^{(i)}$ and $P_n^{(i)}$ using (28) to (30).
  end for
  Scale weights such that $\sum_i w_n^{(i)} = 1$.
end for

Fig. 1. Rao-Blackwellised variable rate particle filter

For each i, initialise particle changepoint sequence with $\{\theta_0^{(i)}\} \leftarrow \{\tau_0^{(i)}, u_0^{(i)}$, where $\tau_0^{(i)} = 0$ and $\{u_0^{(i)}\} \sim p(u_0)$.
for $n = 1 \dots N$ **do**
  for $i = 1 \dots N_F$ **do**
    Sample a history $\theta_{n-1}^{(i)} \sim \sum_j v_{n-1}^{(j)} \delta_{\theta_{n-1}^{(j)}}(\theta_{n-1})$.
    Propose a sequence extension $\theta_{n\backslash n-1}^{(i)} \sim q(\theta_{n\backslash n-1}|\theta_{n-1}^{(i)}, y_n)$.
    Add extension to changepoint sequence $\theta_n^{(i)} \leftarrow \theta_{n-1}^{(i)} \cup \theta_{n\backslash n-1}^{(i)}$.
    Calculate state $\hat{x}_n$ using (12).
    Calculate particle weight using (18).
  end for
  Scale weights such that $\sum_i w_n^{(i)} = 1$.
end for

Fig. 2. Piecewise deterministic variable rate particle filter

B. Conditionally Deterministic Likelihoods

When a conditionally deterministic model is used, the state at observation time t_n is specified by the changepoint sequence θ_n (plus the initial state, x_0), using (12). Thus, the required likelihood term is simply given by,

$$p(y_n|\theta_n, y_{1:n-1}) = p(y_n|\hat{x}_n). \tag{31}$$

This leads to a particle filter for variable rate models with conditionally deterministic dynamics, summarised in Fig. 2. The resulting trajectories $x(t)^{(i)}$ for each particle are realisations from a piecewise-deterministic process, so this algorithm is named the piecewise-deterministic variable rate particle filter (PD-VRPF).

C. Improving the Variable Rate Particle Filter

The bootstrap versions of the VRPFs described in sections III-A and III-B may perform poorly if changepoints are not obvious until significantly after they occur. For example, in a tracking example, if a jump occurs in the acceleration, but only the position is observed, then this change may not be obvious until a number of observations have arrived. In this case, the estimation may be improved by the introduction of resample-move (RM) steps [21]. In an RM scheme, optional Metropolis-Hastings (MH) moves are conducted to alter the particle states after the importance sampling has taken place. For variable rate models, any one of the previous changepoints, τ_k, or associated parameters, u_k, could be adjusted. Because more observations are available than when the changepoint was first proposed, it may be possible to construct more informed proposals and so move the changepoints towards regions with higher posterior probability. It is even possible to retrospectively add or remove changepoints, using reversible jump MH moves [22]. Variable rate particle filters using RM with piecewise deterministic models are described in [3], [23].

Rather than conducting the IS and MH steps separately, it is possible to combine them using the framework of SMC samplers [11]. This was suggested in [3], again for piecewise deterministic dynamics, but the extension to conditionally linear-Gaussian models is straightforward.

Filtering schemes which alter past changepoints – whether using an SMC sampler or RM – are computationally expensive, because many likelihood calculations must be conducted, for each observation from the time of the change onwards. In some cases, it may be simpler to just use a bootstrap filter with more particles.

IV. Variable Rate Particle Smoothing

A filter conducts inference sequentially as new observations are introduced. The purpose of a smoother is to produce a second estimate once all the observations have been made, using future values to improve upon the filter performance. Estimating changepoints online is a challenging task because the presence of a change may not be obvious until after it has happened. Thus, it is anticipated that a smoothing algorithm will provide significantly improved performance at changepoint estimation.

The target distribution for a variable rate smoothing algorithm is the posterior distribution over the entire changepoint sequence, $p(\theta|y_{1:N})$. A particle approximation to this distribution is generated by the final step of the VRPF. However, in the same manner as the fixed rate filter-smoother of [24], this approximation is likely to lack path-space diversity – because of the necessary resampling step in the filtering algorithm, the particles all share the same set of changepoints before a particular time, with variation only appearing for changepoints closer to T. For a good characterisation of the smoothing distribution, it is necessary to rejuvenate the set of particles. This is achieved with a backward pass through the observations in a similar manner to the forward-backward method described in [8].

The target distribution may be factorised as follows,

$$p(\theta|y_{1:N}) = p(\theta_n^+|y_{1:N})p(\theta_n|\theta_n^+, y_{1:N}). \tag{32}$$

This suggests a sequential strategy for smoothing. A particle drawn from $p(\theta_n^+|y_{1:N})$ may be extended backwards by sampling from the backwards conditional distribution, $p(\theta_n|\theta_n^+, y_{1:N})$ and concatenating the two partial sequences. The new particle will be a sample from the target distribution, but the resulting approximation will still suffer from low path-space diversity before time t_{n-1}. Therefore, θ_{n-1} is marginalised by simply discarding the changepoints which occur before this time. This results in a particle drawn from $p(\theta_{n-1}^+|y_{1:N})$. The procedure then continues recursively. After a complete backwards pass through the observations, a single particle from the target distribution will have been generated. The process is then repeated until sufficient samples are obtained.

At step n for a given particle, the future changepoint sequence may be considered to be fixed. It remains to devise a method for drawing samples from the backwards conditional distribution, $p(\theta_n|\theta_n^+, y_{1:N})$. Such methods are discussed in the following sections for the two classes of models.

A. Conditionally Linear-Gaussian Smoothing

When the state dynamics are linear-Gaussian conditional on the changepoint sequence, it is possible split the dependence on the observations into past and future by introducing the current state as an additional variable. This trick was used in [25] in the derivation of the Rao-Blackwellised particle filter for fixed rate conditionally linear-Gaussian models.

$$\begin{aligned}
&p(\theta_n|\theta_n^+, y_{1:N}) \\
&\quad\propto p(\theta_n, \theta_n^+|y_{1:N}) \\
&\quad= \int p(x_n, \theta_n, \theta_n^+|y_{1:N})dx_n \\
&\quad\propto \int p(y_{n+1:N}|x_n, \theta_n, \theta_n^+, y_{1:n})p(x_n, \theta_n, \theta_n^+|y_{1:n})dx_n \\
&\quad= \int p(y_{n+1:N}|x_n, \theta_n^+)p(x_n|\theta_n, y_{1:n})dx_n \\
&\qquad\times p(\theta_n^+|\theta_n)p(\theta_n|y_{1:n})
\end{aligned} \tag{33}$$

Finally, the particle approximation is substituted for the filtering distribution.

$$\hat{p}(\theta_n|\theta_n^+, y_{1:N}) = \sum_i \tilde{w}_n^{(i)} \delta_{\theta_n^{(i)}}(\theta_n) \tag{34}$$

where the backwards conditional weights are given by

$$\begin{aligned}
\tilde{w}_n \propto \int &p(y_{n+1:N}|x_n, \theta_n^+) \\
&\times p(x_n|\theta_n^{(i)}, y_{1:n})dx_n p(\theta_n^+|\theta_n^{(i)})
\end{aligned} \tag{35}$$

As before, normalisation is enforced by scaling the weights so that they sum to 1.

The changepoint transition density may be expressed as,

$$\begin{aligned}
p(\theta_n^+|\theta_n) &= p(\tau_{K_n+1:K}, u_{K_n+1:K}|\tau_{K_n}, u_{K_n}, \tau_{K_n+1} > t_n) \\
&\propto p(u_{K_n+1}|\tau_{K_n+1}, \tau_{K_n}, u_{K_n})p(\tau_{K_n+1}|\tau_{K_n}, u_{K_n})
\end{aligned} \tag{36}$$

The integral in (35) contains two terms involving the current state, x_n. The first is the familiar state posterior generated by the Kalman filter using (24–30).

$$p(x_n|\theta_n^{(i)}, y_{1:n}) = \mathcal{N}(x_n|m_n^{(i)}, P_n^{(i)}) \tag{37}$$

The mean and covariance of this distribution will have been calculated during the filtering stage, and can be stored for use now in the smoother.

The second state term in (35) is the likelihood, $p(y_{n+1:N}|x_n, \tilde{\theta}_n^+)$. This may be considered to be an improper density over x_n, and may be calculated analytically using a backwards Kalman filter, in a similar manner to that used in the two-filter smoother [14], [18], [25]. Such a backwards Kalman filter uses the following recursions. Details are provided in appendix **??**, and in the aforesaid references.

$$p(y_{n+1:N}|x_n, \theta_n^+) \propto \mathcal{N}(x_n|\tilde{m}_n^-, \tilde{P}_n^-) \tag{38}$$

$$p(y_{n:N}|x_n, \theta_n^+) \propto \mathcal{N}(x_n|\tilde{m}_n, \tilde{P}_n) \tag{39}$$

$$\tilde{m}_n^- = A_{n+1}^{-1}\tilde{m}_{n+1} \tag{40}$$

$$\tilde{P}_n^- = A_{n+1}^{-1}(\tilde{P}_{n+1} + Q_{n+1})A_{n+1}^{-T} \tag{41}$$

$$\tilde{\mu}_n = C_n\tilde{m}_n^- \tag{42}$$

$$\tilde{S}_n = C_n\tilde{P}_n^- C_n^T + R_n \tag{43}$$

$$\tilde{K}_n = \tilde{P}_n^- C_n^T \tilde{S}_n^{-1} \tag{44}$$

$$\tilde{m}_n = \tilde{m}_n^- + \tilde{K}_n(y_n - \tilde{\mu}_n) \tag{45}$$

$$\tilde{P}_n = \tilde{P}_n^- - \tilde{K}_n\tilde{S}_n\tilde{K}_n^T \tag{46}$$

Exact methods for initialising this recursion are discussed in appendix **??**. However, it is often sufficient to use an approximation, for example $\tilde{m}_N^{(i)} = m_N^{(i)}$ and $\tilde{P}_N^{(i)} = P_N^{(i)}$.

Substituting into (35), the backwards conditional weights are given by,

$$\tilde{w}_n \propto p(\theta_n^+|\theta_n^{(i)})\mathcal{N}(\tilde{m}_n^-|m_n, \tilde{P}_n^- + P_n). \tag{47}$$

Using these weights, a sample of θ_n may be drawn from the particle distribution of (34), completing the smoothing algorithm. Once sampling has progressed backwards from $n = N \ldots 1$, a particle from the smoothing distribution will have been generated. This procedure may then be repeated until sufficient particles have been obtained. The algorithm is summarised in Fig. 3.

The algorithmic complexity of this variable rate particle smoother for conditionally linear-Gaussian models is $O(N_F \times N_S \times N)$. It is possible to reduce this by using an MCMC sampling scheme in the style of [**?**], in which case it is no longer necessary to calculate the backwards sampling weights for every filter particle.

B. Conditionally Deterministic Smoothing

In a piecewise-deterministic system, the trick of introducing a state variable to split the set of observations into past and future does not work. Furthermore, the simplifications used by the fixed rate particle smoother which exploit the Markovian nature of the state sequence [8] are of no help, because past changepoints are not independent of future observations

1: Run Rao-Blackwellised variable rate particle filter to approximate $p(\theta_n|y_{1:n})$ with weighted particles $\{\theta_n^{(i)}, w_n^{(i)}\}$ and $\{p(x_n|\theta_n^{(i)}, y_{1:n})\}$ as normal distributions with moments $\{m_n^{(i)}\}$ and $\{P_n^{(i)}\}$. Store all results.
2: **for** $i = 1 \dots N_S$ **do**
3: Initialise particle using $\theta^{(i)} \sim \sum_j w_N^{(j)} \delta_{\theta^{(j)}}(\theta)$.
4: Initialise sufficient statistics $\tilde{m}_N^{(i)}$ and $\tilde{P}_N^{(i)}$ (see text).
5: **for** $n = N \dots 1$ **do**
6: Predict state mean and covariance $\tilde{m}_n^{-(i)}$ and $\tilde{P}_n^{-(i)}$ using 40 to 41.
7: **for** $j = 1 \dots N_P$ **do**
8: Calculate weight $\tilde{w}_n^{(j)}$ using (47).
9: **end for**
10: Sample $\theta_n^{(i)} \sim \sum_j \tilde{w}_n^{(j)} \delta_{\theta_n^{(j)}}(\theta_n)$.
11: Update state mean and covariance $\tilde{m}_n^{(i)}$ and $\tilde{P}_n^{(i)}$ using 42 to 46.
12: Discard $\theta_{n-1}^{(i)}$.
13: **end for**
14: **end for**

Fig. 3. Rao-Blackwellised Variable Rate Particle Smoother

(i.e. $p(\theta_n|\theta_n^+, y_{1:N}) \neq p(\theta_n|\theta_n^+, y_{1:n})$). A new approach is required.

Consider the augmented MPP consisting of $\tilde{\theta} = \{\tau_{0:K}, u_{0:K}, x_{0:K}\}$ ($\tilde{\theta}_n$ and $\tilde{\theta}_n^+$ defined in the same manner as before). The mark corresponding to τ_k is now the parameter-state pair (u_k, x_k), i.e. the state at the time of the change plus the parameters for the next division. The smoother is formulated by factorising the augmented sequence posterior distribution as before,

$$p(\tilde{\theta}|y_{1:N}) = p(\tilde{\theta_n}^+|y_{1:N}) p(\tilde{\theta}_n|\tilde{\theta}_n^+, y_{1:N}). \quad (48)$$

The backwards conditional term may be expanded as,

$$p(\tilde{\theta}_n|\tilde{\theta}_n^+, y_{1:N}) \propto p(y_{1:N}|\tilde{\theta}_n, \tilde{\theta}_n^+) p(\tilde{\theta}_n^+|\tilde{\theta}_n) p(\tilde{\theta}_n). \quad (49)$$

For each particle in the filtering approximations, the changepoint state values $x_{0:K_n}$ may be calculated deterministically from θ_n and stored during the filtering procedure. The obvious strategy then is to recursively sample $\tilde{\theta}_n$ from the backwards conditional term, using the re-weighted particles of the nth filtering approximation. However, the conditionally deterministic structure prevents this from working. $x(t)$ is fully determined by $\tilde{\theta}_n$ up to τ_{K_n+1}, but this will not usually match up with x_{K_n+1} as specified in $\tilde{\theta}_n^+$. The result is that $p(\tilde{\theta}_n^+|\tilde{\theta}_n) = 0$ for (almost) all particles. In other words, we are trying to join two deterministic state trajectories together, but they do not meet up in the middle.

The solution to this problem is provided by an idea from the Sequential Monte Carlo (SMC) samplers of [11]. Rather than just using a value of $\tilde{\theta}_n$ from the filtering particles, a replacement, u'_{K_n}, is proposed for the existing parameter corresponding to the final changepoint with a value chosen such that the past and future paths meet up. The modified sequence is denoted $\tilde{\theta}'_n$. Such a proposal may be written as,

$$p(\tilde{\theta}_n|y_{1:n}) q(u_{K_n}|\tilde{\theta}_n, \tilde{\theta}_n^+, y_{1:N}) \propto p(y_{1:n}|\tilde{\theta}_n) p(\tilde{\theta}_n) q(u'_{K_n}|\tilde{\theta}_n, \tilde{\theta}_n^+, y_{1:N}). \quad (50)$$

For this proposal to work, a new condition to be imposed on the state dynamics. For any pair of adjacent changepoints, (τ_k, x_k), and (τ_{k+1}, x_{k+1}) it must be possible to calculate a parameter value which results in the transition from the former to the latter. As a rule of thumb, this requires the number of dimensions of u_k to be greater or equal to the number of dimensions of x_k.

The target distribution is augmented by introduction of an artificial density to cover the new changepoint sequence and the discarded parameter.

$$p(\tilde{\theta}'_n|\tilde{\theta}_n^+, y_{1:N}) L(u_{K_n}|\tilde{\theta}'_n, \tilde{\theta}_n^+, y_{1:N}), \quad (51)$$

It is clear that this new target distribution admits the desired posterior term as a marginal.

The augmented target distribution may be sampled using Markov chain Monte Carlo (MCMC) [26], in a similar manner to the fixed rate smoother of [**?**]. Metropolis-Hastings (MH) steps are conducted to draw samples, $\{\tilde{\theta}_n, u'_{K_n}\}$, from the target distribution by sampling the proposal and accepting the new values with a given probability.

The ratio of target (51) and proposal (50) densities is given by,

$$\begin{aligned} \beta_n \propto & \frac{p(y_{r_n^-:r_n^+}|\tilde{\theta}'_n)}{p(y_{r_n^-:n}|\tilde{\theta}_n)} \\ & \times \frac{p(\tilde{\theta}_n^+|\tilde{\theta}'_n) p(u'_{K_n}|\tilde{\theta}_n \setminus u_{K_n})}{p(u_{K_n}|\tilde{\theta}_n \setminus u_{K_n})} \\ & \times \frac{L(u_{K_n}|\tilde{\theta}_n, \tilde{\theta}_n^+, y_{1:N})}{q(u'_{K_n}|\tilde{\theta}_n, \tilde{\theta}_n^+, y_{1:N})}, \end{aligned} \quad (52)$$

where $r_n^- = \min(m : t_m > \tau_{K_n})$ and $r_n^+ = \max(m : t_m < \tau_{K_n+1})$.

The transition density term will include a delta function with unit probability mass at the point for which the past and future state trajectories match up.

$$\begin{aligned} p(\tilde{\theta}_n^+|\tilde{\theta}'_n) \propto & p(u_{K_n+1}|\tau_{K_n+1}, \tau_{K_n}, u_{K_n}) \\ & \times p(\tau_{K_n+1}|\tau_{K_n}, u_{K_n}) \delta_{f(x_{K_n}, v_{K_n}, \tau_{K_n}, \tau_{K_n+1})}(x_{K_n+1}) \end{aligned} \quad (53)$$

If there is only one value of u'_{K_n} for which the two trajectories match up, then the proposal density $q(u'_{K_n}|\tilde{\theta}_n, \tilde{\theta}_n^+, y_{1:N})$ is also a delta function, with unit probability mass at the one valid value. The remaining terms are likelihoods and parameter transition density terms, defined by the system equations. The simplest choice for $L(.|.)$ is simply as a uniform distribution so that it cancels out in the acceptance probability.

If $\beta_n^{(m-1)}$ is the target-proposal ratio for the current state in the chain, and β_n^* that for the proposed state, then the acceptance probability is given by,

Run variable rate particle filter to approximate $p(\tilde{\theta}_n|y_{1:n})$ with weighted particles $\{\tilde{\theta}_n^{(i)}, w_n^{(i)}\}$. Store all results.
for $i = 1 \dots N_S$ **do**
  Initialise particle from Nth filter approximation $\tilde{\theta}^{(i)} \sim \sum_j w_N^{(j)} \delta_{\tilde{\theta}^{(j)}}(\tilde{\theta})$.
  for $n = N \dots 1$ **do**
    Get the current changepoint history $\tilde{\theta}_n^{(i)(0)} \leftarrow \tilde{\theta}_n^{(i)}$.
    for $m = 1 \dots M$ **do**
      Propose a new changepoint history from nth filter approximation $\tilde{\theta}_n^{(i)*} \sim \sum_j w_n^{(j)} \delta_{\tilde{\theta}_n^{(j)}}(\tilde{\theta}_n)$.
      Propose a replacement parameter $u'^*_{K_n} \sim q(u_{K_n}|\tilde{\theta}_n, \tilde{\theta}_n^+, y_{1:N})$.
      Construct a new changepoint history, $\tilde{\theta}_n'^{(i)*}$, by replacing $u^*_{K_n}$ with $u'^*_{K_n}$.
      Calculate acceptance probability $\alpha_n^{(m)}$ using (52) and (54).
      With probability $\alpha_n^{(m)}$, $\tilde{\theta}_n^{(i)(m)} \leftarrow \tilde{\theta}_n^{(i)*}$. Otherwise, $\tilde{\theta}_n^{(i)(m)} \leftarrow \tilde{\theta}_n^{(i)(m-1)}$.
    end for
    Keep final sample $\tilde{\theta}_n^{(i)} \leftarrow \tilde{\theta}_n^{(i)(M)}$.
  end for
end for

Fig. 4. MCMC piecewise-deterministic variable rate particle smoother

$$\alpha_n^{(m)} = \min\left(1, \frac{\beta_n^*}{\beta_n^{(m-1)}}\right). \tag{54}$$

The algorithm is summarised in Fig. 4.

The Markov chains require no burn-in, because the original value is itself a sample from the target distribution. The number of MH steps, M, at each observation time, n, allows a trade-off of performance against time. Larger values of M will ensure more unique samples in the smoothing approximation, but will also take longer to execute. In the simulations discussed in this report, $M = 1$ was used throughout.

V. Simulations

A. Conditionally Linear-Gaussian Model

The Rao-Blackwellised VRPS algorithm was tested on the financial time series model of [4], [5], in which prices of an asset are treated as noisy observations of a latent state, which evolves according to a drift-diffusion with occasional jumps.

The latent state is a vector with two elements, the underlying value of the asset, and the trend followed by this value.

$$\mathbf{x}(t) = [x(t), \dot{x}(t)]^T \tag{55}$$

This evolves continuously according to a jump-diffusion model:

$$d\mathbf{x}(t) = \begin{bmatrix} 0 & 1 \\ 0 & -\lambda \end{bmatrix} \mathbf{x}(t)dt + \begin{bmatrix} 0 \\ \sigma \end{bmatrix} d\beta(t) + d\mathbf{J}(t) \tag{56}$$

where λ introduces a mean regression effect on the trend and $\beta(t)$ is standard Brownian motion (with unit diffusion constant). The jump term, $d\mathbf{J}(t)$, is zero everywhere except where jumps occur.

$$d\mathbf{J}(t) = \begin{cases} \mathbf{J}_k & t \in \{\tau_k\} \\ 0 & \text{elsewhere} \end{cases} \tag{57}$$

$$\mathbf{J}_k \sim \mathcal{N}(\mathbf{J}_k|\mathbf{0}, Q_{J,u_k}) \tag{58}$$

Jumps occur at a random set of times, $\{\tau_k\}$, and are one of two types: value jumps, indicated by $u_k = 1$, and trend jumps, indicated by $u_k = 2$.

The jump covariance matrices are,

$$Q_{J,u_k} = \begin{cases} \begin{bmatrix} \sigma_{J1}^2 & 0 \\ 0 & 0 \end{bmatrix} & u_k = 1 \\ \begin{bmatrix} 0 & 0 \\ 0 & \sigma_{J2}^2 \end{bmatrix} & u_k = 2 \end{cases}. \tag{59}$$

This model may be discretised at the observation times using standard methods (see e.g. [27]). Assuming observations of value only and Gaussian observation noise with standard deviation σ_y^2, the resulting discrete time dynamics are described by the following equations:

$$\mathbf{x}_n = A\mathbf{x}_{n-1} + \mathbf{w}_n \tag{60}$$

$$y_n = C\mathbf{x}_n + v_n \tag{61}$$

where the $\mathbf{w}_b$ and v_n are Gaussian random variables with covariance matrixes Q_n and R respectively. The time between observations times is denotes $\Delta t = t_n - t_{n-1}$.

$$A = \begin{bmatrix} 1 & \frac{1}{\lambda}(1 - e^{(-\lambda\Delta t)}) \\ 0 & e^{(-\lambda\Delta t)} \end{bmatrix} \tag{62}$$

$$C = \begin{bmatrix} 1 & 0 \end{bmatrix} \tag{63}$$

$$Q_n = \begin{cases} Q_D + Q_{J,u_k} & \exists k : \tau_k \in [t_{n-1}, t_n] \\ Q_D & \text{otherwise} \end{cases} \tag{64}$$

$$Q_D = \frac{\sigma^2}{2\lambda} \begin{bmatrix} q_1 & q_2 \\ q_2 & q_3 \end{bmatrix} \tag{65}$$

$$q_1 = \frac{1}{\lambda^2}(2\lambda\Delta t - (3 - e^{(-\lambda\Delta t)})(1 - e^{(-\lambda\Delta t)})\ ) \tag{66}$$

$$q_2 = \frac{1}{\lambda}(1 - e^{(-\lambda\Delta t)})^2 \tag{67}$$

$$q_3 = 1 - e^{(-2\lambda\Delta t)} \tag{68}$$

$$R = [\sigma_y^2] \tag{69}$$

The times between changepoints were assumed to be exponentially distributed, with equal probability of value and trend jumps.

$$p(\tau_k, u_k|\tau_{k-1}, u_{k-1}) = P(u_k)p(\tau_k|\tau_{k-1}) \tag{70}$$

$$p(\tau_k|\tau_{k-1}) = \begin{cases} \frac{1}{\alpha}\exp(\alpha(\tau_k - \tau_{k-1})) & \tau_k > \tau_{k-1} \\ 0 & \tau_k < \tau_{k-1} \end{cases} \tag{71}$$

$$P(u_k) = \begin{cases} 0.5 & u_k = 1 \\ 0.5 & u_k = 2 \end{cases} \tag{72}$$

The algorithms were first tested on artificial data simulated from this model. The following parameters were used: $\Delta t =$

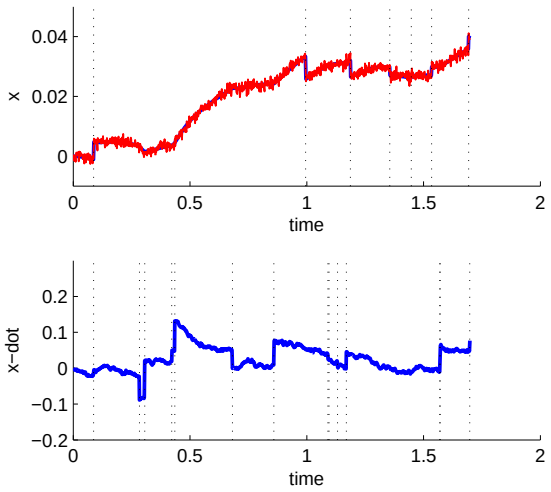


Fig. 5. An example simulated data set, showing value (top) and trend (bottom). Value observations are overlayed. Jump times are shown as dotted vertical lines.

0.0017, $N = 1000$, $\alpha = 20$, $\lambda = 5$, $\sigma = 0.05$, $\sigma_{J1} = 0.005$, $\sigma_{J2} = 0.05$, $\sigma_y = 0.001$.

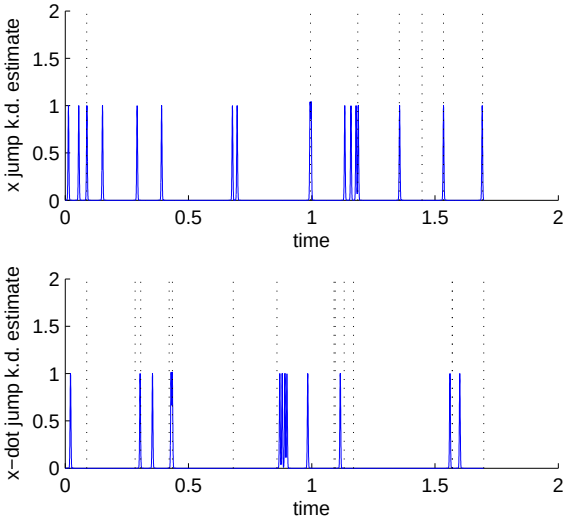
The filter used $N_P = 100$ particles, and the smoother resampled $N_S = 100$ sequences. Bootstrap proposals were used for the filter.

An example realisation simulated from the model is shown in figure 5.

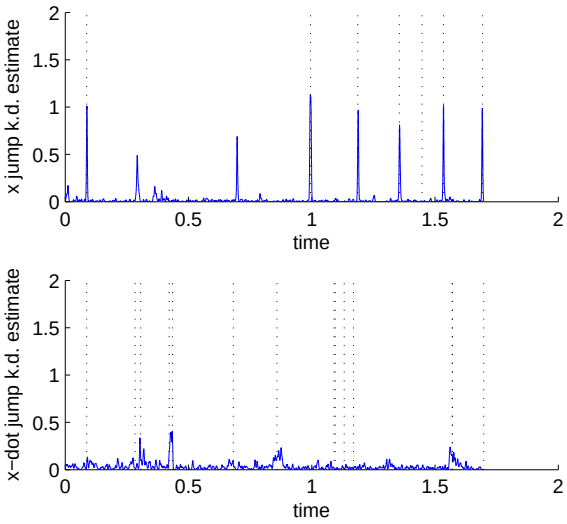
Both the smoother and the final processing step of the filter produce a particle approximation to the changepoint sequence posterior, $p(\theta|y_{1:N})$. It is actually no simple task to compare the quality of these approximations. Each particle consists of a list of changepoint times and types, and there is no obvious way to take an average over the set to produce a single estimate. Furthermore, once a best estimate has been calculated, it is not obvious how the error should be assessed, as it is not trivial to define a distance metric between lists of varying lengths.

For a first, qualitative comparison of the changepoint sequence distributions, it is possible to generate kernel density estimates for the changepoint times. These are constructed by simply adding a (unit amplitude) Gaussian kernel for each changepoint present in each particle and scaling by the number of particles. The resulting function provides an approximate measure of the likelihood of finding a jump at a given time. These are shown in figure 6 for the example data set.

It is immediately apparent that the approximation produced by the filter lacks diversity. The jump time kernel density estimate comprises peaks with magnitude 1, with regions of 0 between. This is a result of every particle containing an almost identical list of changepoints, as a result of resampling. In contrast, the smoother is clearly providing a less degenerate particle representation, and generally contains larger kernel density peaks at the times of larger jumps. Furthermore, the kernel density for the trend jump times contains broader peaks, capturing the fact that because the trend is not observed directly, the jumps are harder to localise precisely.



(a)



(b)

Fig. 6. Filter (a) and smoother (b) kernel density estimates for value (top) and trend (bottom) jump times. Correct times overlayed as dotted vertical lines.

A further comparison of the algorithms is possible via the state estimates they generate. Here we can compare three options: the filtering results, using the VRPF and a Kalman filter for the state estimates; the filter-smoother results, using the final VRPF approximation for the changepoint sequence and a Rauch-Tung-Striebel (RTS) smoother for the state estimates; and the full smoothing results, using the VRPS followed by an RTS smoother. For the example data set, the trend estimates are shown in figure 7 (The estimates of value are less informative, as this quantity is observed). Again, the improved particle diversity of the smoother is apparent.

For a quantitative comparison, the algorithms were tested on 10 realisations from the model, each of 1000 time steps, and the following statistics were calculated for each:

- The number of unique changepoint sequences. This is a

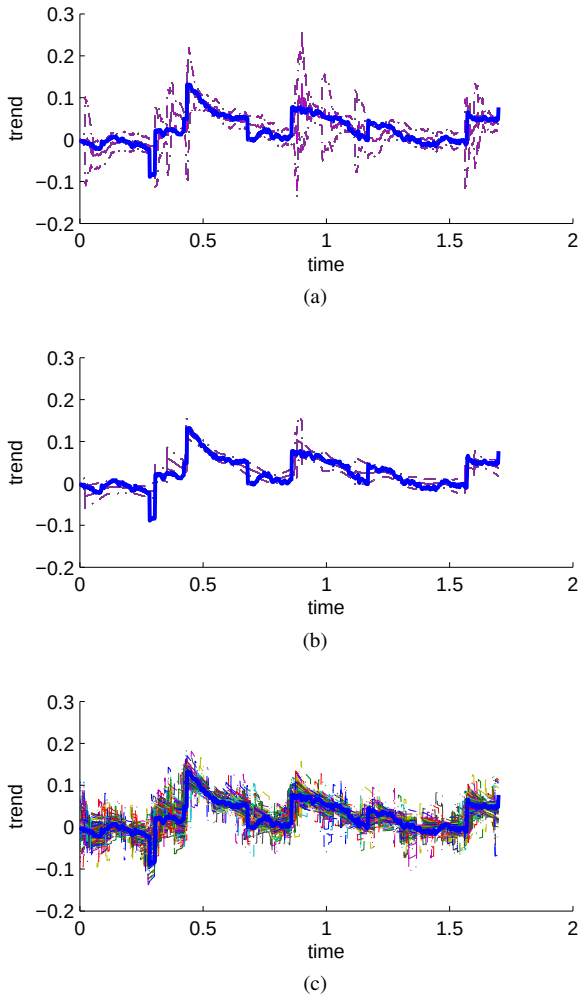


Fig. 7. Trend estimates using RBVRPF + Kalman filter (a) RBVRPF + RTS smoother (b) and RBVRPS + RTS smoother. Solid lines show means of each particle. Dashed lines show mean ± 2 standard deviations.

TABLE I
CHANGEPOINT SEQUENCE ESTIMATION PERFORMANCE.

	RBVRPF	RBVRPS
mean unique no. sequences	7.8	100
mean unique no. jump times	47.8	1076.7

measure of particle diversity of the approximation.

- The number of unique changepoint times. Another measure of particle diversity.
- The root-mean-square error (RMSE) of the mean state estimate. The mean state estimate is the average of the Gaussian means from all the particles.

The results are shown in tables I and II.

The new RBVRPS algorithm outperforms the filter in terms of both state estimation accuracy and particle diversity.

Finally, we demonstrate the algorithms on a real financial data set, 1000 data points representing 1.7s of USD-GBP foreign exchange rate data from April 2008, using the same parameters as for the previous simulations. The data shown in Fig. 8, along with jump time kernel density estimates from the filter and smoother. In this example, there is no ground truth against which to judge the results. However, the smoother does a pleasing job of estimating jump times in the same areas that one would if analysing the data "by eye". We submit this simply to illustrate the practical applications of the model and algorithm.

TABLE II
STATE ESTIMATION PERFORMANCE.

	RBVRPF and KF	RBVRPF and RTS	RBVRPS and RTS
mean value estimate RMSE	5.31×10^{-4}	4.48×10^{-4}	4.16×10^{-4}
mean trend estimate RMSE	2.56×10^{-2}	1.79×10^{-2}	1.49×10^{-2}

REFERENCES

[1] M. Jacobsen, *Point process theory and applications: marked point and piecewise deterministic processes*, C. Newman and S. Resnick, Eds. Birkhauser, 2006.

[2] S. Godsill, J. Vermaak, W. Ng, and J. Li, "Models and algorithms for tracking of maneuvering objects using variable rate particle filters," *Proceedings of the IEEE*, vol. 95, no. 5, pp. 925 –952, May 2007.

[3] N. Whiteley, A. M. Johansen, and S. Godsill, "Monte carlo filtering of piecewise deterministic processes," *Journal of Computational and Graphical Statistics*, vol. 20, no. 1, pp. 119–139, 2011. [Online]. Available: http://pubs.amstat.org/doi/abs/10.1198/jcgs.2009.08052

[4] S. Godsill, "Particle filters for continuous-time jump models in tracking applications," *ESAIM: Proc.*, vol. 19, pp. 39–52, 2007. [Online]. Available: http://dx.doi.org/10.1051/proc:071907

[5] H. L. Christensen, J. Murphy, and S. J. Godsill, "Forecasting high-frequency futures returns using online langevin dynamics," *submitted to IEEE journal of selected topics in signal processing*, vol. under review, p. under review, 2012.

[6] N. Gordon, D. Salmond, and A. Smith, "Novel approach to nonlinear/non-Gaussian Bayesian state estimation," *Radar and Signal Processing, IEE Proceedings F*, vol. 140, no. 2, pp. 107 –113, April 1993.

[7] A. Doucet, S. Godsill, and C. Andrieu, "On sequential Monte Carlo sampling methods for Bayesian filtering," *Statistics and Computing*, vol. 10, pp. 197–208, 2000, 10.1023/A:1008935410038. [Online]. Available: http://dx.doi.org/10.1023/A:1008935410038

[8] S. J. Godsill, A. Doucet, and M. West, "Monte Carlo smoothing for nonlinear time series," *Journal of the American Statistical Association*, vol. 99, no. 465, pp. 156–168, 2004. [Online]. Available: http://pubs.amstat.org/doi/abs/10.1198/016214504000000151

[9] O. Cappé, S. Godsill, and E. Moulines, "An overview of existing methods and recent advances in sequential Monte Carlo," *Proceedings of the IEEE*, vol. 95, no. 5, pp. 899–924, 2007.

[10] A. Doucet and A. M. Johansen, "A tutorial on particle filtering and smoothing: Fifteen years later," in *The Oxford Handbook of Nonlinear Filtering*, D. Crisan and B. Rozovsky, Eds. Oxford University Press, 2009. [Online]. Available: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.157.772&rep=rep1&type=pdf

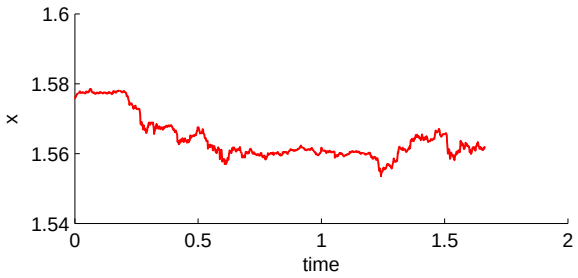
[11] P. Del Moral, A. Doucet, and A. Jasra, "Sequential Monte Carlo samplers," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 68, no. 3, pp. 411–436, 2006.

[12] R. E. Kalman, "A new approach to linear filtering and prediction problems," *Journal Of Basic Engineering*, vol. 82, no. Series D, pp. 35–45, 1960. [Online]. Available: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.129.6247&rep=rep1&type=pdf

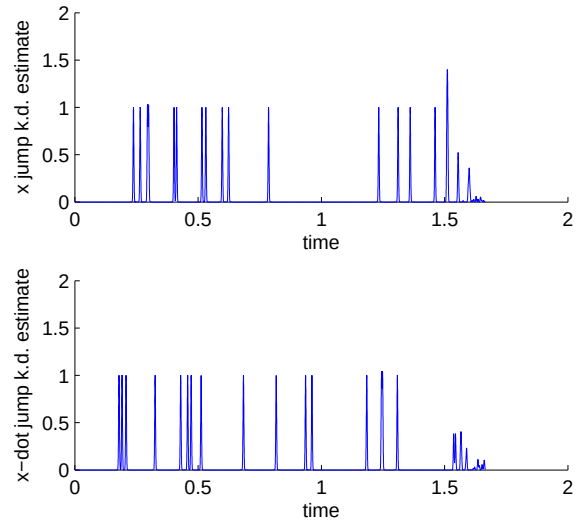
[13] H. Rauch, F. Tung, and C. Striebel, "Maximum likelihood estimates of linear dynamic systems," *AIAA journal*, vol. 3, no. 8, pp. 1445–1450, 1965.

[14] D. Fraser and J. Potter, "The optimum linear smoother as a combination of two optimum linear filters," *Automatic Control, IEEE Transactions on*, vol. 14, no. 4, pp. 387 – 390, aug 1969.

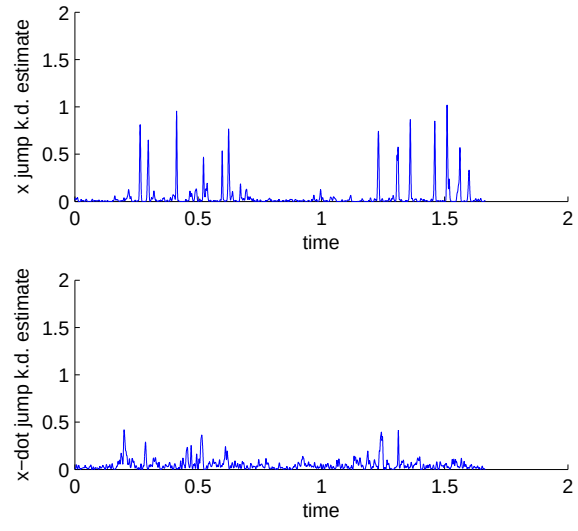
[15] X. Li and V. Jilkov, "Survey of maneuvering target tracking. part i: Dynamic models," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 39, no. 4, pp. 1333–1364, 2003.



(a)



(b)



(c)

Fig. 8. Foreign exchange data (a), along with filter (a) and smoother (b) kernel density estimates for the state and trend jump times.

[16] M. K. Pitt and N. Shephard, "Filtering via simulation: Auxiliary particle filters," *Journal of the American Statistical Association*, vol. 94, no. 446, pp. pp. 590–599, 1999. [Online]. Available: http://www.jstor.org/stable/2670179

[17] R. Douc, O. Cappe, and E. Moulines, "Comparison of resampling schemes for particle filtering," in *Image and Signal Processing and Analysis, 2005. ISPA 2005. Proceedings of the 4th International Symposium on*, sept. 2005, pp. 64 – 69.

[18] B. Anderson and J. Moore, *Optimal filtering*. Prentice-Hall (Englewood Cliffs, NJ), 1979.

[19] G. Casella and C. P. Robert, "Rao-blackwellisation of sampling schemes," *Biometrika*, vol. 83, no. 1, pp. 81–94, 1996. [Online]. Available: http://biomet.oxfordjournals.org/content/83/1/81.abstract

[20] A. Doucet, N. De Freitas, K. Murphy, and S. Russell, "Rao-Blackwellised particle filtering for dynamic Bayesian networks," in *Proceedings of the Sixteenth Conference on Uncertainty in Artificial Intelligence*. Citeseer, 2000, pp. 176–183.

[21] W. R. Gilks and C. Berzuini, "Following a moving target-Monte Carlo inference for dynamic Bayesian models," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 63, no. 1, pp. 127–146, 2001. [Online]. Available: http://dx.doi.org/10.1111/1467-9868.00280

[22] P. Green, "Reversible jump Markov chain Monte Carlo computation and bayesian model determination," *Biometrika*, vol. 82, no. 4, p. 711, 1995.

[23] K. Gilholm and N. O. Everett, "A novel algorithm for tracking targets using manoeuvre models," in *Target Tracking and Data Fusion: Algorithms and Applications, 2008 IET Seminar on*, april 2008, pp. 143 –148.

[24] G. Kitagawa, "Monte Carlo filter and smoother for non-Gaussian nonlinear state space models," *Journal of Computational and Graphical Statistics*, vol. 5, no. 1, pp. pp. 1–25, 1996. [Online]. Available: http://www.jstor.org/stable/1390750

[25] S. Sarkka, P. Bunch, and S. Godsill, "A backward-simulation based rao-blackwellized particle smoother for conditionally linear gaussian models," in *Preprint submitted to 16th IFAC Symposium on System Identification.*, 2012.

[26] W. Gilks, S. Richardson, and D. Spiegelhalter, *Markov chain Monte Carlo in practice*. Chapman and Hall, 1996.

[27] M. S. Grewal and A. P. Andrews, *Kalman Filtering: Theory and Practice Using MATLAB*. John Wiley & Sons, Inc., 2002. [Online]. Available: http://dx.doi.org/10.1002/0471266388