The Rao-Blackwellised Variable Rate Particle Smoother for Conditionally Linear-Gaussian Models

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

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

***Index Terms*—**

I. Introduction

IN model-based schemes for probabilistic inference, some unknown quantity is treated as a random process which evolves over time according to a dynamic model. This latent state is observed at a discrete set of times via another random process modelling the measurement mechanism. Using the two models and applying Bayes rule, inference of the hidden state can be made from the observations.

For simple models with linear dynamics and Gaussian-distributed random variables, optimal analytic inference algorithms exist, including the Kalman filter [1] and Rauch-Tung-Striebel (RTS) smoother [2]. For nonlinear, non-Gaussian models, such analytic solutions do not exist, and it is often necessary to employ numerical approximations, including the particle filter [3] and particle smoother [4], [5] (see [6], [7] for a thorough introduction to particle methods).

Commonly, the unknown quantity under consideration is continually varying – e.g. the position of a moving object – but is modelled as a discrete-time random process synchronous with the observations. This leads to the standard discrete-time hidden Markov model. Such "fixed rate" models are poorly suited to quantities with discontinuities in their evolution. 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 jumps in the acceleration when turns begin or end. In such cases, a "variable rate" model may be more appropriate, in which the state evolution is dependent upon a set of unknown changepoints.

The discontinuous nature of variable rate models makes them inherently nonlinear. However, in some cases, some subset of the state may behave according to linear-Gaussian dynamics when conditioned on the set of random changepoints and remaining nonlinear components of the state. Such models have been introduced in [8], [9], along with an algorithm for conducting sequential inference of both the changepoints and state. This uses the method of Rao-Blackwellisation (see, e.g. [10], [11]) to exploit the conditionally linear-Gaussian structure of the model. A particle filter is used to estimate the distribution of the changepoint times and nonlinear state variables, after which a Kalman filter is used to estimate the linear state for each particle. This is the Rao-Blackwellised variable rate particle filter (RBVRPF).

In this paper, a new algorithm is described for use with conditioanlly linear-Gaussian variable rate models for estimation of the smoothing distribution, i.e. the distribution over the sequence of states given all the observations. This uses a similar derivation to that of the fixed rate Rao-Blackwellised particle smoother (RBPS) of [12]. The new algorithm is called the Rao-Blackwellised variable rate particle smoother (RBVRPS).

We review the structure of conditionally linear-Gaussian variable rate models in section II and revise the RBVRPF in section III, using a new notation to clarify the derivations. The new smoothing algorithm is presented in section IV with supporting simulations in section V.

II. Conditionally Linear-Gaussian Variable Rate Models

The notation associated with variable rate models can become convoluted because the observations and changepoints are not concurrent.

We consider a general model from time 0 to T, between which observations, $\{y_1 \dots y_N\}$, are made at times $\{t_1 \dots t_N\}$. The linear state at these times is written as $\{x_1 \dots x_N\}$. During this period, an unknown number, K, of changepoints occur at times $\{\tau_1 \dots \tau_K\}$, each with an associated changepoint parameters, $\{u_1 \dots u_K\}$. Discrete sets containing multiple values over time will be written as, e.g. $y_{1:n} = \{y_1 \dots y_n\}$. To keep the following derivations concise, we introduce an additional variable for each changepoint-parameter pair $\theta_k = \{\tau_k, u_k\}$.

It will be desirable to denote the set of changepoints and parameters (of unspecified size) which occur within some range of time. This will be written as $\theta_{[s,t]} = \{\theta_k \forall k : s < \theta_k < t\}$. Note that such a variable not only conveys where changepoints occur, but also where they do not within the specfied range.

A conditionally linear-Gaussian variable rate model can now be expressed by the following system equations:

$$\theta_k \sim p(\theta_k|\theta_{1:k-1}) \tag{1}$$

$$x_n = A(\theta_{[0,t_n]})x_{n-1} + w_n \tag{2}$$

$$y_n = C(\theta_{[0,t_n]})x_n + v_n. \tag{3}$$

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.

The random variables w_n and v_n have a zero-mean Gaussian distribution with covariance matrices Q_n and R_n respectively.

Thus, if $\theta_{[0,T]}$ is given, the filtering distributions, $p(x_n|\theta_{[0,t_n]}, y_{1:n})$, and smoothing distributions, $p(x_n|\theta_{[0,T]}, y_{1:N})$, can be calculated optimally using a Kalman filtering and smoothing methods (see e.g. [13]). It remains to use a particle filter or smoother to estimate $\theta_{[0,T]}$.

III. The Rao-Blackwellised Variable Rate Particle Filter

The Rao-Blackwellised Variable Rate Particle Filter (RBVRPF) was first described in [8], and in [9] it was developed for use in a financial prediction algorithm.

The objective of the algorithm is to sequentially estimate the distribution of the variable rate componentes, $p(\theta_{[0,t_n]}|y_{1:n})$, at each time t_n. The linear state filtering distribution, $p(x_n|\theta_{[0,t_n]}, y_{1:n})$, can then be estimated by a Kalman filter.

The target distribution may be expanded using Bayes theorem.

$$\begin{aligned} p(\theta_{[0,t_n]}|y_{1:n}) &\propto P(y_n|\theta_{[0,t_n]}, y_{1:n-1}) p(\theta_{[0,t_n]}|y_{1:n-1}) \\ &= p(y_n|\theta_{[0,t_n]}, y_{1:n-1}) p(\theta_{[t_{n-1},t_n]}|\theta_{[0,t_{n-1}]}) p(\theta_{[0,t_{n-1}]}|y_{1:n-1}) \end{aligned} \tag{4}$$

This distribution cannot be calculated analytically due to the combinatorial complexity of the possible numbers, positions and parameters of changepoints. Instead, a particle filter can be used to approximate the distribution numerically.

A particle filter is an algorithm used to approximate a filtering density with a discrete set of weighted samples drawn from that density using sequential importance sampling. For details, see e.g. [6], [7]. In this case, each particle will consist of list of changepoint times between 0 and t.

$$\hat{p}(\theta_{[0,t_n]}|y_{1:n}) = \sum_i w_n^{(i)} \delta_{\theta^{(i)}_{[0,t_n]}}(\theta_{[0,t_n]}) \tag{5}$$

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

The particle filter is a recursive algorithm. At the $(n)th$ step, a particle, $\{\theta^{(i)}_{[0,t_{n-1}]}\}$, is first resampled from those approximating the filtering distribution at the $(n-1)th$ step, using an arbitrary set of proposal weights.

$$q(\theta^{(i)}_{[0,t_{n-1}]}|y_{1:n-1}) = \sum_i v_{n-1}^{(i)} \delta_{\theta^{(i)}_{[0,t_{n-1}]}}(\theta_{[0,t_{n-1}]}) \tag{6}$$

An extension, $\{\theta_{[t_{n-1},t_n]}\}$, is then proposed from an importance distribution, $q(\theta_{[t_{n-1},t_n]}|\theta_{[0,t_{n-1}]})$. Finally, the particle is weighted according to the ratio of the target distribution and the proposal.

$$\begin{aligned} w_n^{(i)} &= \frac{p(\theta^{(i)}_{[0,t_n]}|y_{1:n})}{q(\theta^{(i)}_{[0,t_{n-1}]}|y_{1:n-1}) q(\theta^{(i)}_{[t_{n-1},t_n]}|\theta^{(i)}_{[0,t_{n-1}]})} \\ &\propto \frac{p(y_n|\theta^{(i)}_{[0,t_n]}, y_{1:n-1}) p(\theta^{(i)}_{[t_{n-1},t_n]}|\theta^{(i)}_{[0,t_{n-1}]}) p(\theta^{(i)}_{[0,t_{n-1}]}|y_{1:n-1})}{q(\theta^{(i)}_{[0,t_{n-1}]}|y_{1:n-1}) q(\theta^{(i)}_{[t_{n-1},t_n]}|\theta^{(i)}_{[0,t_{n-1}]})} \\ &\approx \frac{w_{n-1}^{(i)}}{v_{n-1}^{(i)}} \times \frac{p(y_n|\theta^{(i)}_{[0,t_n]}, y_{1:n-1}) p(\theta^{(i)}_{[t_{n-1},t_n]}|\theta^{(i)}_{[0,t_{n-1}]})}{q(\theta_{[t_{n-1},t_n]}|\theta_{[0,t_{n-1}]})} \end{aligned} \tag{7}$$

The normalisation may be enforced by scaling the weights so that they sum to 1. This is the RBVRPF of [8], [9].

Now consider terms of 7 in turn. First, $p(y_n|\theta_{[0,t_n]}, y_{1:n-1})$ is the predictive likelihood estimated by the Kalman filter, which will be a Gaussian distributed.

$$p(y_n|\theta_{[0,t_n]}, y_{1:n-1}) = \mathcal{N}(y_n|\mu_n, S_n) \tag{8}$$

The mean and variance are given by the following standard recursions (dependence on $\theta_{[0,t_n]}$ suppressed for clarity).

$$m_n^- = A_n m_{n-1} \tag{9}$$

$$P_n^- = A_n P_n A_n^T + Q_n \tag{10}$$

$$\mu_n = C_n m_n^- \tag{11}$$

$$S_n = C_n P_n^- C_n^T + R_n \tag{12}$$

$$K_n = P_n^- C_n^T S_n^{-1} \tag{13}$$

$$m_n = m_n^- + K_n(y_n - \mu_n) \tag{14}$$

$$P_n = P_n^- - K_n S_n K_n^T \tag{15}$$

Next consider the transition term $p(\theta_{[t_{n-1},t_n]}|\theta_{[0,t_{n-1}]})$. Technically, any number of new changepoints could occur in the interval $[t_{n-1}, t_n]$. If k changepoints have occured before t_{n-1}, then the probability of a particuar set of changepoints within this interval is given by:

$$\begin{aligned} &p(\theta_{[t_{n-1},t_n]}|\theta_{[0,t_{n-1}]}) \\ &\quad = \prod_{j=1}^{J} p(\theta_{k+j}|\theta_{1:k+j-1}, \tau_{k+1} > t_{n-1}) p(\tau_{k+J+1} > t_n|\theta_{1:k+J}, \tau_{k+1} > \end{aligned}$$

where J is the number of changepoints occuring in the interval. Practically, the probability of J being larger than 1 is small; if changepoints were occurring this often then there would be little point in using a variable rate model! Thus, there are two significant cases: either a new changepoint occurs in the interval $[t_{n-1}, t_n]$ or it does not. The transition probability then simplifies to:

$$\begin{aligned} p(\theta_{[t_{n-1},t_n]}|\theta_{[0,t_{n-1}]}) &= \begin{cases} p(\theta_{k+1}|\theta_{1:k}, \tau_{k+1} > t_{n-1}) p(\tau_{k+2} > t_n|\theta_{1:k+1}) \\ p(\tau_{k+1} > t_n|\theta_{1:k}, \tau_{k+1} > t_{n-1}) \end{cases} \\ &= \begin{cases} \frac{p(\theta_{k+1}|\theta_{1:k}) p(\tau_{k+2} > t_n|\theta_{1:k+1})}{p(\tau_{k+1} > t_{n-1}|\theta_{1:k})} & \tau_{k+1} < t_n \\ \frac{p(\tau_{k+1} > t_n|\theta_{1:k})}{p(\tau_{k+1} > t_{n-1}|\theta_{1:k})} & \tau_{k+1} < t_n \end{cases} \end{aligned}$$

This can be calculated from transition model for τ_k and u_k specified by 1. For the most basic "bootstrap" [3] form of RBVRPF, this transition density may be used as the importance distribution, leading to the usual simplification in the weight formula.

The choice of proposal weights, $\{v_{n-1}^{(i)}\}$, requires particular attention in the design of RBVRPFs. 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. A particle pays a debt in transition probability when a changepoint is proposed, and does not see it repaid in likelihood until later. 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 [14], in which proposal weights are given by:

$$v_{n-1}^{(i)} \propto \max(1, N_F w_{n-1}^{(i)}) \tag{19}$$

where N_F is the number of filtering particles. The RBVRPF is summarised below.

Set initial particle sufficient statistics, $\{m_0^{(i)}\}$ and $\{P_0^{(i)}\}$ according to prior.
Initialise particles $\{\theta_{[0,0]}^{(i)}\} \leftarrow \emptyset$.
for n=1 ...N **do**
  for i=1 ...M **do**
    Sample a changepoint history $\theta_{[0,t_{n-1}]}^{(i)} \sim \sum_j v_{n-1}^{(j)} \delta_{\theta_{[0,t_{n-1}]}^{(j)}}(\theta_{[0,t_{n-1}]})$.
    Propose $\theta_{[t_{n-1},t_n]}^{(i)} \sim q(\theta_{[t_{n-1},t_n]}|\theta_{[0,t_{n-1}]}^{(i)}, y_{1:n})$
    Calculate $\mu_n^{(i)}$ and $S_n^{(i)}$ using (9) to (12).
    Calculate new state mean and covariance $m_n^{(i)}$ and $P_n^{(i)}$ using (13) to (15).
    Weight $w_n^{(i)} \propto \frac{w_{n-1}^{(i)}}{v_{n-1}^{(i)}} \times \frac{\mathcal{N}(y_n|\mu_n^{(i)}, S_n^{(i)}) p(\theta_{[t_{n-1},t_n]}^{(i)}|\theta_{[0,t_{n-1}]}^{(i)})}{q(\theta_{[t_{n-1},t_n]}|\theta_{[0,t_{n-1}]})}$
  end for
  Scale weights such that $\sum_i w_n^{(i)} = 1$
end for

IV. The Rao-Blackwellised Variable Rate Particle Smoother

Estimating changepoints online is a challenging task because the presence of a change may not be obvious until after it has happened. It is thus expected that a smoothing algorithm will provide significantly improved performance at changepoint estimation. In this section, the same Rao-Blackwellisation method is used to develop a particle smoother for variable rate models with linear-Gaussian state dynamics (RBVRPS). The derivation follows a similar course to that for the fixed rate Rao-Blackwellised Particle Smoother of [12].

The particles of the final filtering step approximate the distribution, $p(\theta_{[0,T]}|y_{1:N})$, which is desired smoothing distribution. However, in the same manner as the fixed rate filter-smoother of [15], this approximation is likely to be degenerate – the particles all share the same set of changepoints from early times, with variation only appearing for changepoints close to T. For a good characterisation of the smoothing distribution, it is necessary to rejeuventate the set of particles. This is achieved with a backward pass through the observations in a similar manner to the methods described in [5], [12].

The target smoothing distribution may be expanded with Bayes rule.

$$p(\theta_{[0,T]}|y_{1:N}) = p(\theta_{[t_n,T]}|y_{1:N}) p(\theta_{[0,t_n]}|\theta_{[t_n,T]}, y_{1:N}) \tag{20}$$

Thus, a set of particles representing $p(\theta_{[t_n,T]}|y_{1:N})$ may be extended backwards by sampling from the backwards conditional distribution, $p(\theta_{[0,t_n]}|\theta_{[t_n,T]}, y_{1:N})$, which may be approximated by reweighting the partilce of the $(n)th$ filtering distribution. The resulting particles are then marginalised by discarding the changepoints which come before t_{n-1} (which will still be suffering from low particle diversity), and the procedure continues recursively.

If the future changepoints and their parameters, $\tilde{\theta}_{[t_n,T]}$, have already been sampled (and may thus be considered fixed), then the backward conditional term may be expressed in terms of the filtering distribution.

$$\begin{aligned}
&p(\theta_{[0,t_n]}|\tilde{\theta}_{[t_n,T]}, y_{1:N}) \\
&\quad \propto p(\theta_{[0,t_n]}, \tilde{\theta}_{[t_n,T]}|y_{1:N}) \\
&\quad = \int p(x_n, \theta_{[0,t_n]}, \tilde{\theta}_{[t_n,T]}|y_{1:N}) dx_n \\
&\quad \propto \int p(y_{n+1:N}|x_n, \theta_{[0,t_n]}, \tilde{\theta}_{[t_n,T]}, y_{1:n}) p(x_n, \theta_{[0,t_n]}, \tilde{\theta}_{[t_n,T]}|y_{1:n}) dx_n \\
&\quad = \int p(y_{n+1:N}|x_n, \theta_{[0,t_n]}, \tilde{\theta}_{[t_n,T]}) p(x_n|\theta_{[0,t_n]}, y_{1:n}) dx_n p(\tilde{\theta}_{[t_n,T]}|\theta_{[0,t
\end{aligned}$$

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

$$\begin{aligned}
&p(\theta_{[0,t_n]}|\tilde{\theta}_{[t_n,T]}, y_{1:N}) \\
&\quad \propto \sum_i W_n^{(i)} \int p(y_{n+1:N}|x_n, \theta_{[0,t_n]}^{(i)}, \tilde{\theta}_{[t_n,T]}) p(x_n|\theta_{[0,t_n]}^{(i)}, y_{1:n}) dx_n p(\tilde{\theta} \\
&\quad = \sum_i \tilde{w}_n^{(i)} \delta_{\theta_{[0,t_n]}^{(i)}}(\theta_{[0,t_n]})
\end{aligned}$$

where the backwards conditional weights are given by

$$\tilde{w}_n \propto \int p(y_{n+1:N}|x_n, \theta_{[0,t_n]}^{(i)}, \tilde{\theta}_{[t_n,T]}) p(x_n|\theta_{[0,t_n]}^{(i)}, y_{1:n}) dx_n p(\tilde{\theta}_{[t_n,T]}|\theta_{[0,t_n]}^{(i)} \quad (23$$

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

If k changepoints occur before time t_n, then the changepoint transition term $p(\tilde{\theta}_{[t_n,T]}|\theta_{[0,t_n]}^{(i)})$ may be expressed as:

$$p(\tilde{\theta}_{[t_n,T]}|\theta_{[0,t_n]}^{(i)}) = p(\tilde{\theta}_{k+1:K}|\theta_{1:k}, \tau_{k+1} > t_n) \tag{24}$$

For a Markovian sequence of changepoints, this is simply proportional to the probability of the first changepoint after t_n given the last changepoint preceeding t_n, i.e.

$$P(\tilde{\theta}_{[t_n,T]}|\theta^{(i)}_{[0,t_n]}) \propto P(\tilde{\theta}_{k+1}|\theta_k, \tau_{k+1} > t_n) \tag{25}$$

The second term in equation 23 is the familiar Kalman filter distribution $P(x_n|\theta^{(i)}_{[0,t_n]}, y_{1:n}) = \mathcal{N}(x_n|m^{(i)}_n, P^{(i)}_n)$. The first term, $p(y_{n+1:N}|x_n, \theta^{(i)}_{[0,t_n]}, \tilde{\theta}_{[t_n,T]})$, is an improper likelihood density, and may be calculated analytically using a backwards Kalman filter, in a similar manner to that used in the two-filter smoother [12], [13], [16]. Such a backwards Kalman filter uses the following recursions. Details are provided in **??**, and in the aforesaid references.

$$p(y_{n+1:N}|x_n, \theta^{(i)}_{[0,t_n]}, \tilde{\theta}_{[t_n,T]}) = Z_n \mathcal{N}(x_n|\tilde{m}_n, \tilde{P}_n) \tag{26}$$

$$\tilde{m}^-_n = A^{-1}_{n+1}\tilde{m}_{n+1} \tag{27}$$
$$\tilde{P}^-_n = A^{-1}_{n+1}(\tilde{P}_{n+1} + Q_{n+1})A^{-T}_{n+1} \tag{28}$$
$$\tilde{\mu}_n = C_n \tilde{m}^-_n \tag{29}$$
$$\tilde{S}_n = C_n \tilde{P}^-_n C^T_n + R_n \tag{30}$$
$$\tilde{K}_n = \tilde{P}^-_n C^T_n \tilde{S}^{-1}_n \tag{31}$$
$$\tilde{m}_n = \tilde{m}^-_n + \tilde{K}_n(y_n - \tilde{\mu}_n) \tag{32}$$
$$\tilde{P}_n = \tilde{P}^-_n - \tilde{K}_n \tilde{S}_n \tilde{K}^T_n \tag{33}$$

Substituting into equation 23, the backwards conditional weights are given by:

$$\tilde{w}_n \propto p(\theta_{[t_n,T]}|\theta^{(i)}_{[0,t_n]})\mathcal{N}(\tilde{m}^-_n|m_n, \tilde{P}^-_n + P_n) \tag{34}$$

Samples of $\theta_{[0,t_n]}$ may be drawn from this particle distribution. Once sampling has progressed backwards from $n = N \dots 1$, a complete particle from the smoothing distribution has been generated. This procedure may then be repeated until sufficient particles have been obtained. The procedure is summarised below.

Run Rao-Blackwellised particle filter to approximate $p(\theta_{[0,t_n]}|y_{1:n})$ with particles $\theta^{(i)}_{[0,t_n]}$ and associated Gaussian moments $m^{(i)}_n$ and $P^{(i)}_n$.
for $i = 1 \dots N_S$ **do**
  for $n = N \dots 1$ **do**
    Backwards Kalman filter: Calculate $\tilde{m}^{-(i)}_n$ and $\tilde{P}^{-(i)}_n$ using 27 to 33.
    for $j = 1 \dots N_P$ **do**
      $\tilde{w}^{(j)}_n \propto w^{(j)}_n p(\theta_{[t_n,T]}|\theta^{(j)}_{[0,t_n]})\mathcal{N}(\tilde{m}^{-(j)}_n|m_n, \tilde{P}^{-(j)}_n + P_n)$
    end for
    Sample $\tilde{\theta}^{(i)}_{[0,t_n]} \sim \sum_j \tilde{w}^{(j)}_n \delta_{\theta^{(j)}_{[0,t_n]}}(\theta_{[0,t_n]})$
    Discard $\tilde{\theta}^{(i)}_{[0,t_{n-1}]}$
  end for
end for

V. Simulations

The RBVRPS algorithm was tested on the financial time series model of [8], [9], in which prices of a 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.

$$\mathbf{x}_n = [x_n, \dot{x}_n]^T \tag{35}$$

This evolves continuously according to a drift-diffusion model:

$$d\mathbf{x}_t = \begin{bmatrix} 0 & 1 \\ 0 & -\lambda \end{bmatrix} \mathbf{x}_t dt + \begin{bmatrix} 0 \\ \sigma \end{bmatrix} d\beta(t) \tag{36}$$

where λ introduces a mean regression effect on the trend and $\beta(t)$ is standard brownian motion (with unit diffusion constant).

In addition, the state experiences jumps at random times, $\{\tau_k\}$. Two types of jumps occur: value jumps, indicated by $u_k = 1$, and trend jumps, indicated by $u_k = 2$. The magnitudes of jumps are zero-mean Gaussian-distributed with standard deviation σ_{J1} and σ_{J2} respectively. This model may be discretised at the observation times by matrix fraction decomposition (see e.g. [17]*Not sure what to cite here. Simo's thesis? Or the references therein which I've never seen?*). Assuming Gaussian observation noise with standard deviation σ^2_y, the resulting discrete time dynamics are described by the following equations (see **??**):

$$\mathbf{x}_n = A\mathbf{x}_{n-1} + \mathbf{w}_n \tag{37}$$
$$y_n = C\mathbf{x}_n + v_n \tag{38}$$

where the $\mathbf{w}_b$ and v_n are Gaussian random variables with covariance matrixes Q_n and R resepectively.

$$A = \begin{bmatrix} 1 & \frac{1}{\lambda}(1 - e^{(-\lambda T)}) \\ 0 & e^{(-\lambda T)} \end{bmatrix}$$

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

$$Q_n = \begin{cases} Q_{\text{diff}} + Q_{\text{jump}} & \exists k : \tau_k \in [t_{n-1}, t_n] \\ Q_{\text{diff}} & \text{otherwise} \end{cases}$$

$$Q_{\text{jump}} = \begin{cases} \begin{bmatrix} \sigma_J 1^2 & 0 \\ 0 & 0 \end{bmatrix} & u_k = 1 \\ \begin{bmatrix} 0 & 0 \\ 0 & \sigma_J 2^2 \end{bmatrix} & u_k = 2 \end{cases}$$

$$Q_{\text{diff}} = \frac{\sigma^2}{2\lambda} \begin{bmatrix} \frac{1}{\lambda^2}(2\lambda T - (3 - e^{(-\lambda T)})(1 - e^{(-\lambda T)}) & \frac{1}{\lambda}(1 - e^{(-\lambda T)})^2 \\ \frac{1}{\lambda}(1 - e^{(-\lambda T)})^2 & 1 - e^{(-2\lambda T)} \end{bmatrix}$$

$$R = [\sigma^2_y]$$

This model fits nicely into the Rao-Blackwellised variable rate filtering and smoothing schemes.

The algorithms were first tested on artificial data simulated from the model.

VI. Conclusion

The conclusion goes here.

Appendix A
Derivation of the Backwards Kalman Filter

Appendix one text goes here.

Appendix B
Model Discretisation

Appendix two text goes here.

Acknowledgment

The authors would like to thank...

References

[1] 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

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

[3] 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, apr 1993.

[4] 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

[5] 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

[6] 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.

[7] 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

[8] 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

[9] 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.

[10] 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

[11] 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.

[12] 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.

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

[14] 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.

[15] 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

[16] 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.

[17] S. Sarkka, "Recursive bayesian inference on stochastic differential equations," Ph.D. dissertation, Helsinki University of Technology Laboratory of Computational Engineering, 2006.

Pete Bunch Biography text here.

Simon Godsill Biography text here.