MODELS AND ALGORITHMS FOR TRACKING USING VARIABLE DIMENSION PARTICLE FILTERS

Simon Godsill and Jaco Vermaak

Signal Processing Group
Department of Engineering, University of Cambridge, U.K.
sjg@eng.cam.ac.uk, jv211@eng.cam.ac.uk

ABSTRACT

In this paper we discuss modifications to tracking models, and sequential Monte Carlo algorithms for their estimation from sequential and batch data. New models for tracking are proposed which involve a dynamical model on both the hidden state value and its arrival times. In this way we aim to have a more flexible and parsimonious representation of time-varying state characteristics which is more amenable to estimation using Bayesian filtering. In order to perform inference in this scenario new particle filters and smoothers are proposed for cases where the state process arrives at unknown times that are generally different from the observation arrival times.

1. INTRODUCTION

In tracking problems it is often required to adapt the parameters of the model automatically to the characteristics of the target. In particular, targets wishing to evade detection may adopt unpredictable behaviour, exhibiting both fast and slow manoeuvres, see for example figure 1 in which we see both sharp turns and long straight sections. In classical filtering methods it is typically hard to find approaches robust enough to track the variable characteristics of a target with time. In model-based filtering it is possible to capture the time-varying nature of a target if the models are sufficiently flexible. However, this can lead to relatively complex models that might be seen as overparameterised.

Here we propose a departure from the standard state-space modelling approach to the problem in which the models are re-expressed in terms of a random time-arrival process for new states, which is in general different from the (known) arrival rate of the data. In this way we hope to be able to track rapid manoeuvres with many state values, while smoother trajectories are modelled with only a few state values - see figure 1. A dynamic model is thus specified for both the time arrivals of new states and the state values at those times. Since the observed data are measured on a different time grid from the states, the likelihood function for data given states is specified in a special way which depends on states in a local time region surrounding each data point. A simple example of this involves a local polynomial or spline fit through the state values, with likelihoods evaluated at the interpolated values of the state process corresponding to the arrival time of each data point.

The models are necessarily of variable dimension, since it is not known *a priori* how many state values will occur within a given time interval. We discuss new sequential Monte Carlo ('particle filtering') methods [1–3] for inference in this modified state-space framework, including both forward filters for the purely on-line setting and backward smoothers for retrospective inference about a target trajectory.

By expressing the state process in a more parsimonious manner, we aim to achieve more efficient algorithms with less degeneracy than particle filters applied to standard state-space models. In a related problem from audio processing an early version of the models, with state process on a fixed grid of time values, has already shown some significant improvements in estimation accuracy and reduction of degeneracy over the standard approaches [4, 5]. In applications to kernel regression [6, 7] the new sequential variable dimension methods have shown promise compared to standard batch-based approaches.

2. VARIABLE-DIMENSION STATE-SPACE MODELS

The modified state-space model is specified in terms of a random state arrival process:

$$\tau_k \sim f_1(\tau_k|\tau_{k-1})$$

and corresponding parameter values:

$$\theta_k \sim f_2(\theta_k|\theta_{k-1}, \tau_k, \tau_{k-1})$$

Or, with $x_k = [\theta_k, \tau_k]$,

$$x_k \sim f(x_k|x_{k-1})$$

Then, a likelihood model must be specified for the observations. However, this is not immediate since data points are on a different time scale to observations. Each observation y_t is now assumed to depend on a local neighbourhood $\mathcal{N}_t$ of θ_k and τ_k values:

$$y_t \sim g(y_t|\{\theta_k, \tau_k;\, k \in \mathcal{N}_t\})$$

The model is illustrated in figure 2, in which a neighbourhood $\mathcal{N}_t$ of two adjacent states is assumed. One of the

The work of both authors is partially supported by QinetiQ Contract CU006 0000014890 under the Weapon, Platform and Effectors domain of the MOD UK research programme

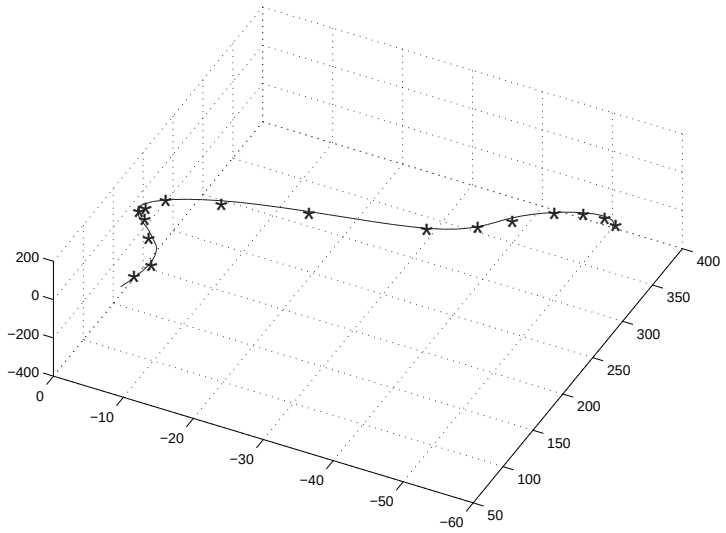

Fig. 1. Example: non-uniform time-sampling.

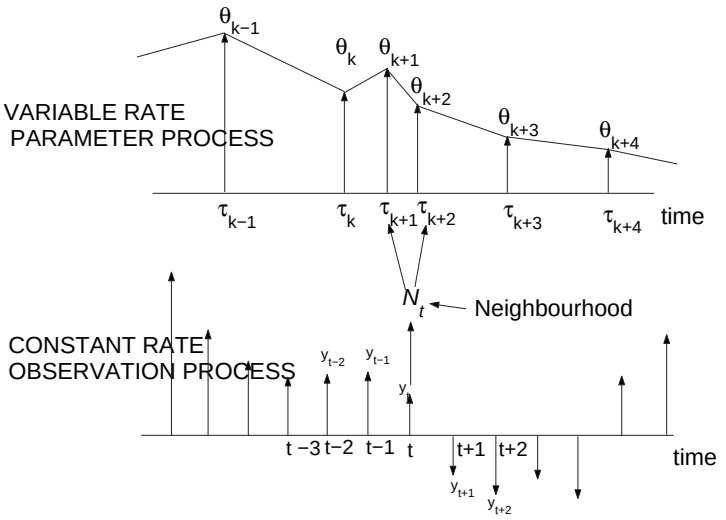

Fig. 2. Variable rate parameter process.

simplest realisations of this process is a linear interpolation in which the two closest τ_k values either side of t are used to interpolate an intermediate state value, so that $\mathcal{N}_t = \{\tau_k, \tau_{k-1}\}$ and

$$\phi_t = \theta_k + \frac{t - \tau_{k-1}}{\tau_k - \tau_{k-1}}(\theta_k - \theta_{k-1})$$

Here τ_k and θ_k are respectively the k-th abscissa and ordinate values. Then, for example, we can choose $g(y_t|\theta_{\mathcal{N}_t}) = g(y_t|\phi_t) = \mathcal{N}(\phi_t, \sigma_t)$, where $\mathcal{N}()$ is the normal distribution. In more sophisticated settings splines or other smooth regression functions can be adopted for the interpolation, leading to a more complex neighbourhood $\mathcal{N}_t$.

2.1. Filtering for the variable-dimension case

As for standard state-space models, Bayesian filtering algorithms can be based on recursions which update the posterior distribution at time t to that at time $t+1$. A convenient way to express this recursion in the non-uniform time-sampling case is in terms of a 'long' sequence of states $x_{1:K}$, where K is considered to be countable but beyond the time horizon of the current time point t. This artifice is helpful in the derivation in that it maintains a fixed-dimension target distribution, for which standard importance sampling ideas are readily understood. An alternative derivation involves a dual-space idea to extend the parameter space to constant dimensionality [6, 7].

First consider the posterior distribution at time $t-1$ for the time sequence of states $x_{1:K}$:

$$\begin{aligned} p(x_{1:K}|y_{1:t-1}) &= p(x_{1:\max(\mathcal{N}_{t-1})}|y_{1:t-1}) \\ &\quad \times f(x_{\max(\mathcal{N}_{t-1})+1:K}|x_{\max(\mathcal{N}_{t-1})}) \end{aligned}$$

and its update to time t:

$$\begin{aligned} p(x_{1:K}|y_{1:t}) \propto & p(x_{1:\max(\mathcal{N}_{t-1})}|y_{1:t-1}) \\ & \times f(x_{\max(\mathcal{N}_{t-1})+1:\max(\mathcal{N}_t)}|x_{\max(\mathcal{N}_{t-1})}) \\ & \times p(y_t|x_{\mathcal{N}_t}) \times f(x_{\max(\mathcal{N}_t)+1:K}|x_{\max(\mathcal{N}_t)}) \end{aligned}$$

Here, $\max(\mathcal{N}_t)$ denotes the member of $\mathcal{N}_t$ having the largest time index. Note also that the terms of the general form '$f(x_{k+1:k+l}|x_k)$' in these expressions indicate the state transition probability for a collection of adjacent states, which can be computed from the single time state transition densities as follows:

$$f(x_{k+1:k+l}|x_k) = \prod_{j=k+1}^{k+l} f(x_j|x_{j-1})$$

The ratio of these posterior distributions is then given by

$$\rho_t = \frac{p(x_{1:K}|y_{1:t})}{p(x_{1:K}|y_{1:t-1})} \propto g(y_t|x_{\mathcal{N}_t})$$

since

$$\begin{aligned} & f(x_{\max(\mathcal{N}_{t-1})+1:K}|x_{\max(\mathcal{N}_{t-1})}) \\ & \quad = f(x_{\max(\mathcal{N}_{t-1})+1:\max(\mathcal{N}_t)}|x_{\max(\mathcal{N}_{t-1})}) \\ & \quad \times f(x_{\max(\mathcal{N}_t)+1:K}|x_{\max(\mathcal{N}_{t-1})}) \end{aligned}$$

This simple result, which parallels that for the uniform time-sampling case, leads to a practical 'bootstrap' style algorithm [1] for Monte Carlo filtering in the non-uniform time-sampling case. The principle of it is as follows.

- Assume that at time t we have a collection of state sequences, or 'particles' $\{x^{(i)}_{1:K}\}_{i=1}^{N}$ each drawn from the posterior distribution $p(x_{1:K}|y_{1:t})$. Since K is beyond the feasible time horizon of states at time t, it is possible to evaluate the likelihood function $g(y_t|x^{(i)}_{\mathcal{N}_t})$ from a valid neighbourhood of states $x^{(i)}_{\mathcal{N}_t}$ in each particle.
- Now consider updating the particles to time $t+1$ with the introduction of a new data point y_{t+1}. Consider $p(x_{1:K}|y_{1:t})$, from which we have N (approximate) draws $x^{(i)}_{1:K}$, to be an importance function for the new posterior distribution $p(x_{1:K}|y_{1:t+1})$. The importance weight for the ith particle is thus given by:

$$w^{(i)}_{t+1} = \frac{p(x^{(i)}_{1:K}|y_{1:t+1})}{p(x^{(i)}_{1:K}|y_{1:t})} \propto g(y_{t+1}|x_{\mathcal{N}_{t+1}})$$

- Notice however that the importance weight does not depend on any states outside of the neighbourhood $\mathcal{N}_{t+1}$. This is the key to a practical implementation scheme, since states beyond $\mathcal{N}_{t+1}$ need not be generated at all: we simply need to maintain sufficient state values at each time t to complete a valid neighbourhood $\mathcal{N}_t$ for the computation of the likelihood $g(y_t|x_{\mathcal{N}_t})$.
- As for the standard particle filter, the scheme may be modified to incorporate an alternative importance function [2, 3], say $q(x_t|x_{t-1})$, which may also depend on the observed data. The derivation for the variable data-rate particle filter can be modified fairly straightforwardly to incorporate this case, and the importance weights then become:

$$w^{(i)}_{t+1} \propto \frac{g(y_{t+1}|x_{\mathcal{N}_{t+1}}) f(x_{\max(\mathcal{N}_t)+1:\max(\mathcal{N}_{t+1})}|x_{\max(\mathcal{N}_t)})}{q(x_{\max(\mathcal{N}_t)+1:\max(\mathcal{N}_{t+1})}|x_{\max(\mathcal{N}_t)})}$$

- We have considered particles which are unweighted at time t, i.e. resampling is assumed at every time step. This would not make sense for many variable rate models and we would recommend resampling much more infrequently. In this case the weights are accumulated as for the standard particle filter at time steps where resampling has not occurred.

Thus a simple filtering scheme can be summarised as follows:

1. Initialize a collection of particles at time $t = 0$. The procedure involves drawing from the initial distribution $f(x_0)$ and then the transition distribution $f(x_1|x_0)$, $f(x_2|x_1)$ etc. until each particle contains a valid neighbourhood of time indices $\mathcal{N}_0$. Set the initial weights to $w^{(i)}_t = 1/N$.
2. Then, for $t = 1$ to ..., and $i = 1$ to N

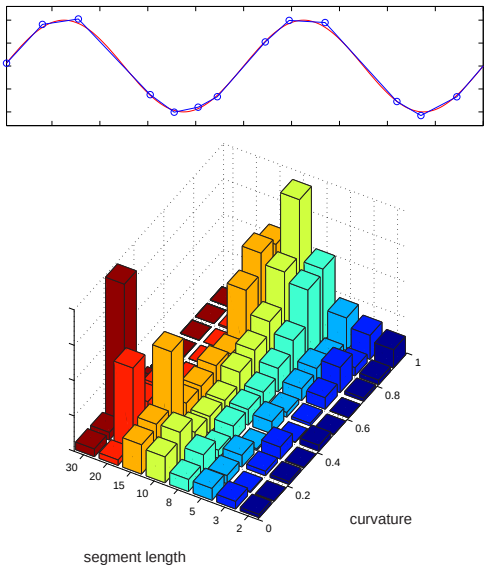

Fig. 3. Example: linear interpolation in 1D. (Top) Sinus curve and linear interpolator. (Bottom) Empirical segment length-curvature distribution.

- Sample

$$x^{(i)}_{\max(\mathcal{N}_{t-1})+1:\max(\mathcal{N}_t)} \sim f(x_{\max(\mathcal{N}_{t-1})+1:\max(\mathcal{N}_t)}|x^{(i)}_{\max(\mathcal{N}_{t-1})})$$

  Important note: this step involves drawing a random number of states x_k, continuing until a valid neighbourhood $\mathcal{N}^{(i)}_t$ is obtained for time t.
- Calculate weight:

$$w^{(i)}_t \propto w^{(i)}_{t-1} \times p(y_t|x^{(i)}_{\mathcal{N}_t})$$

- (Optional) - resample particles with replacement according to weights $w^{(i)}_t$, setting $w^{(i)}_t = 1/N$.

To conclude this section we show some results in figures 3 and 4 of applying the filtering algorithm to the problem of linear interpolation in one and two dimensions, respectively. More knot points are allocated to areas of high curvature, whereas areas of low curvature are associated with longer segments. This is further exemplified by the empirical segment length-curvature distribution in the bottom of figure 3 that clearly shows that short segments are relatively infrequent and mostly associated with high curvature regions, while in contrast longer segments occur in higher frequency, and are almost uniquely confined to regions of low curvature.

2.2. Monte Carlo Smoothing

A Monte Carlo backwards smoothing procedure similar in principle to [8–10] can be devised for the variable dimension problem. Although the forward filtering procedure generates in principle a smoothed set of state trajectories, the

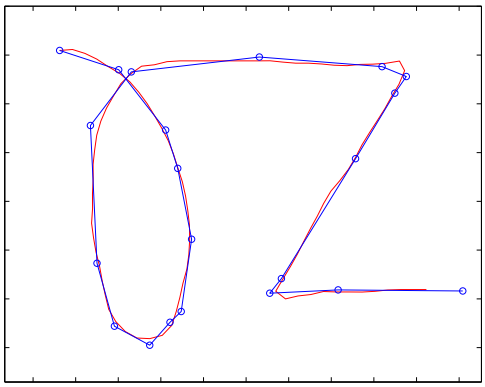

Fig. 4. Example: linear interpolation in 2D.

inherent degeneracy in forward filtering procedures can be reduced by backwards smoothing. In essence we are able to generate new trajectories that allow cross-over between the trajectories stored in the filtering pass. Specifically, we need to store, for each time t, the local neighbourhood of state values in each particle, $\{x_{\mathcal{N}_t}^{(i)}\} \sim p(x_{\mathcal{N}_t}|y_{1:t})$. The idea proceeds as follows. We can write the distribution of a state from time just prior to $\mathcal{N}_{t+1}$ conditional on future states and all data $y_{1:T}$ as:

$$\begin{aligned} & p(x_{\min(\mathcal{N}_{t+1})-1}|x_{\min(\mathcal{N}_{t+1}):K}, y_{1:T}) \\ & \quad \propto p(x_{\min(\mathcal{N}_{t+1})-1}|y_{1:t}) \\ & \qquad \times p(x_{\mathcal{N}_{t+1}}|x_{\min(\mathcal{N}_{t+1})-1}) p(y_{t+1}|x_{\mathcal{N}_{t+1}}) \\ & \quad \propto p(x_{\min(\mathcal{N}_{t+1})-1}|y_{1:t}) \\ & \qquad \times p(x_{\min(\mathcal{N}_{t+1})}|x_{\min(\mathcal{N}_{t+1})-1}) \end{aligned}$$

This indicates how to generate smoothed trajectories based on the stored 'filtering' densities $p(x_{\mathcal{N}_t}|y_{1:t})$, as follows,

- Sample $\tilde{x}_{\mathcal{N}_T}$ from the final output of the particle filtering distribution at time T
- For $t = T-1$ to 1:
  - Choose $\tilde{x}_{\mathcal{N}_t} = x_{\mathcal{N}_t}^{(i)}$ with probability

$$\propto p(\tilde{x}_{\min(\mathcal{N}_{t+1})}|x_{\min(\mathcal{N}_{t+1})-1}^{(i)})$$

- Finally, the sampled sequence $\tilde{x}_{1:\max(\mathcal{N}_T)}$ is a random sample from $p(x_{1:\max(\mathcal{N}_T)}|y_{1:T})$

3. CONCLUSION

We have presented models and algorithms for tracking in environments where targets can be expected to have a high degree of mobility and time-varying characteristics. New particle filtering and smoothing algorithms have been proposed for Monte Carlo inference in such settings. Further work will present detailed simulation in realistic tracking settings, compared with the rather artificial settings presented here.

4. REFERENCES

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

[2] A. Doucet, S. J. Godsill, and C. Andrieu, "On sequential Monte Carlo sampling methods for Bayesian filtering," *Statistics and Computing*, vol. 10, pp. 197–208, 2000.

[3] J S Liu and R Chen, "Sequential Monte Carlo methods for dynamical systems," *J. Am. Statist. Assoc.*, vol. 93, pp. 1032–44, 1998.

[4] W. Fong and S. J. Godsill, "Sequential Monte Carlo simulation of dynamical models with slowly varying parameters: application to audio," in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing*, 2002.

[5] W. Fong and S. J. Godsill, "Sequential Monte Carlo simulation of dynamical models with slowly varying parameters: An extension," in *XI European Signal Processing Conference (EUSIPCO)*, 2002.

[6] J. Vermaak, S. J. Godsill, and A. Doucet, "Radial basis function regression using trans-dimensional sequential monte carlo," in *IEEE Workshop on Statistical Signal Processing*, 2003.

[7] J. Vermaak, S. J. Godsill, and A. Doucet, "Sequential bayesian kernel regression," in *Advances in Neural Information Processing Systems 16, Cambridge, MA*. 2003, MIT Press, (to appear).

[8] A. Doucet, S. J. Godsill, and M. West, "Monte Carlo filtering and smoothing with application to time-varying spectral estimation," in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing*, 2000, vol. II, pp. 701–704, ISBN 0-7803-6296-9.

[9] W. Fong, S. J. Godsill, A. Doucet, and M. West, "Monte Carlo smoothing with application to speech enhancement," *IEEE Trans. on Signal Processing*, vol. 50, no. 2, pp. 438–449, Feb. 2002, Special issue on Monte Carlo Methods.

[10] S. J. Godsill, A Doucet, and M West, "Monte Carlo smoothing for non-linear time series," *J. Am. Stat. Soc.*, 2003, In Press.