Fixed Lag Particle Filtering for Target Tracking

Pete Bunch

August 2011

A report on the first year's progress

UNIVERSITY OF
CAMBRIDGE

Contents

1

Introduction

Everyone loves to track things. Yay!

2

Particle Methods

2.1 Bayes of our lives - what is filtering?

Many tasks in signal processing, science in general, and indeed life, require us to make some estimate of an unknown quantity from indirect, incomplete, or inaccurate observations. By constructing a model to explain how these observations depend on the underlying state, we can infer something about that state. We will express this observation model in terms of a likelihood function:

$$P(Y|X) \tag{2.1}$$

where X is the state and Y the observations. This is not the whole story - in many cases we are not estimating our unknown state "from scratch". Previous experience, prejudice, and prior knowledge can also contribute to our estimates. The likelihood and prior terms can be combined through our friend, Bayes rule (Bayes & Price 1763, **?**), to calculate the posterior probability of the state, i.e. the probability of the state given the observations:

$$P(X|Y) = \frac{P(Y|X)P(X)}{P(Y)} \tag{2.2}$$

This is the basis of the process of inference. Mathematically, we can assign a probability distribution to the state space of X. By applying Bayes rule, we are updating our belief about the values of X using the information in Y.

Often the quantity in which we are interested, X, is changing over time, and we would like to estimate its value at each point in time given only the observations received so far. In this case we will generally assume that the state is Markovian, i.e. that $X_t|X_{t-1}$ is independent of $X_{1:t-2}$, giving us a hidden Markov state-space model. This is the traditional filtering problem. By constucting a model for the evolution of the unknown state, we can now derive our prior information from our estimate at the previous time step. In discrete time, we now write:

$$P(X_t|Y_{1:t}) = \frac{\int P(Y_t|X_t)P(X_t|X_{t-1})P(X_{t-1}|Y_{1:t-1})dX_{t-1}}{P(Y_t|Y_{1:t-1})} \tag{2.3}$$

where the subscript indicates the time and ranges are notated by : in the MATLAB style. $P(X_t|Y_{1:t})$ is called the filtering distribution. Equation 2.3 describes the ideal Bayesian filter.

Instead of marginalising the previous state, we may sometimes want to consider the joint state distribution over all time instances. This may similarly be expanded as:

$$P(X_{1:t}|Y_{1:t}) = \frac{P(Y_t|X_t)P(X_t|X_{t-1})P(X_{1:t-1}|Y_{1:t-1})}{P(Y_t|Y_{1:t-1})} \tag{2.4}$$

The filtering distribution may then be obtained by marginalising out the previous states.

So far, we have expressed the problem entirely in terms of distributions. The same models may be expressed in terms of difference equations of random variables. In the most general form:

$$X_t = f_t(X_{t-1}, V_t) \tag{2.5}$$

$$Y_t = g_t(X_t, W_t) \tag{2.6}$$

where f_t and g_t are known deterministic functions and V_t and W_t and random variables, known as the process and observation noise respectively.

2.2 Keep Kalman carry on - the Kalman filter and its extensions

2.2.1 *The basic filter*

In a few simple cases the filtering set-up permits the derivation of closed form posterior distributions at each time instant. Most notably, the Kalman filter (KF) (Kalman 1960) is an analytic filter for models with continuous state and observation variables, in which both transition and observation models are linear transformations with Gaussian innovations.

$$x_t = Ax_t + v_t \tag{2.7}$$

$$y_t = Cx_t + w_t \tag{2.8}$$

where v_t and w_t are now Gaussian random variables with zero mean and covariance matrices Q and R. We use lower case variables here to emphasise that these are "nice", continuous vectors. (As we shall see, our state variable will later be sets or lists).

Kalman's solution for the linear-Gaussian case is given by:

$$P(x_t|y_{1:t}) = \mathcal{N}(x_t|\mu_t, \Sigma_t) \tag{2.9}$$

$$P(x_t|y_{1:t-1}) = \mathcal{N}(x_t|\hat{\mu}_t, \hat{\Sigma}_t) \tag{2.10}$$

where μ_t, Σ_t, etc. are given by the following recursions.

Time Updates:

$$\hat{\mu}_t = A\mu_{t-1} \tag{2.11}$$

$$\hat{\Sigma}_t = A\Sigma_{t-1}A^T + Q \tag{2.12}$$

Measurement Updates:

$$z_t = y_t - C\hat{\mu}_t \tag{2.13}$$

$$S_t = C\hat{\Sigma}_t C^T + R \tag{2.14}$$

$$K_t = \hat{\Sigma}_t C^T S_t^{-1} \tag{2.15}$$

$$\mu_t = \hat{\mu}_t + K_t z_t \tag{2.16}$$

$$\Sigma_t = (I - K_t C)\hat{\Sigma}_t \tag{2.17}$$

The KF is delightful because it not only provides us with a closed-form analytic solution, but the complexity of that solution does not increase as we receive additional measurements. This is a consequence of the fact that the Gaussian distribution is its own conjugate prior. Unfortunately, no other such convenient cases have been discovered (Daum 2005). Analytic solutions to non-linear, non-Gaussian filtering problems generally require unacceptable conditions, such as zero process noise $Q = 0$ (Daum 2005).

2.2.2 The extended filter

Given the loveliness of the KF, the instinct when faced by an intractable non-linear filtering problem is to linearise it. This produces the Extended Kalman Filter (EKF), and is achieved by replacing the A and C matrices in equations 2.11 through 2.17 above with Jacobians:

$$A_t = \left.\frac{\partial f}{\partial x_t}\right|_{\mu_{t-1}} \tag{2.18}$$

$$C_t = \left.\frac{\partial g}{\partial x_t}\right|_{\hat{\mu}_t} \tag{2.19}$$

2.2.3 The Kalman smoother

The KF gives us an optimum estimate of $P(x_t|y_{1:t})$. However, once more data has arrived, we can improve this estimate. For a given set of data, $y_{1:T}$, we can estimate the optimum estimates for all previous state distributions, $P(x_{1:T}|y_{1:T})$ using a Rauch-Tung-Striebel (RTS) smoother, (Rauch, Tung & Striebel 1965). This begins with a normal KF, followed by a backward filtering pass which propagates information to earlier time instances. This backward pass is implement by the following recursions:

$$\tilde{\mu}_t = \mu_t + \Sigma_t A^T \hat{\Sigma}_{t+1}^{-1}(\tilde{\mu}_{t+1} - \hat{\mu}_{t+1}) \tag{2.20}$$

$$\tilde{\Sigma}_t = \Sigma_t + [\Sigma_t A^T \hat{\Sigma}_{t+1}^{-1}](\tilde{\Sigma}_{t+1} - \hat{\Sigma}_{t+1})[\Sigma_t A^T \hat{\Sigma}_{t+1}^{-1}]^T \tag{2.21}$$

giving us

$$P(x_t|Y_{1:T}) = \mathcal{N}(x_t|\tilde{\mu}_t, \tilde{\Sigma}_t) \tag{2.22}$$

For a full derivation, see (Rauch et al. 1965). There exist other ways to implement Kalman smoothing in a fixed-interval sense, such as the forward-backward smoother, and in a fixed-lag sense, but they will not be used in this work.

2.3 Tough as old bootstraps - a review of particle filter

In general we will not be so lucky as to have a problem with linear-Gaussian dynamics. In this case, a particle filter (PF) may be the best alternative. The PF is based on the idea that if we cannot represent a probability distribution analytically then we may approximate it as a collection of discrete samples or "particles" drawn from the distribution. Often, it will not even be possible to sample from the desired distribution. We therefore use importance sampling (IS) or Markov chain Monte Carlo methods to generate these samples. The principle of a particle filter is that we recursively generate such a set of particles to approximate $P(X_{1:t}|Y_{1:t})$ given a previous set representing $P(X_{1:t-1}|Y_{1:t-1})$.

The particle filter

2.4 Particular details - mathematical basics of the particle filter

With a PF, we approximate a probability distribution with a set of (weighted) samples drawn from that distribution.

$$P(X) \approx \frac{1}{N} \sum_m W^{(m)} \delta_{X^{(m)}}(X) \tag{2.23}$$

where $\delta(x)$ represents a unit probability point mass at a point x, and $\sum_m W^{(m)} = 1$. Such a method allows us to represent any probability distribution of arbitrary complexity, including multidimensional, multimodal, mixed distributions. As the number of particles increases, the accuracy of the approximation improves at the expense of computational complexity. Thus we have the required tool for estimation in non-linear, non-Gaussian scenarios.

We still face the problem of how to generate these samples. The conventional method for this is Sequential Importance Sampling with Resampling (SISR). For a more complete and traditional introduction to this method, see (Cappé, Godsill & Moulines 2007) or (Doucet & Johansen 2009). Here we follow an outline similar to that used for the derivation of the auxiliary particle filter of (Pitt & Shephard 1999).

Suppose we have a particle approximation to the joint posterior distribution from the previous frame, $\hat{P}(X_{1:t-1}|Y_{1:t-1})$. Each particle represents a path through time, $X^{(m)}_{1:t-1}$, and has an associated weight, $W^{(m)}_{t-1}$. Let us imagine also that the unweighted particles are samples from another distribution:

$$\mu(X_{1:t-1}|Y_{1:t-1}) \approx \frac{1}{N} \sum_m W^{(m)} \delta_{X^{(m)}_{1:t}}(X_{1:t}) \tag{2.24}$$

Thus for a given particle,

$$\hat{P}(X^{(m)}_{1:t-1}|Y_{1:t-1}) = W^{(m)}_t \mu(X^{(m)}_{1:t-1}|Y_{1:t-1}) \tag{2.25}$$

We would like to generate a new particle set which includes the current time instance. We propose a new set of extended tracks from a factored proposal distribution $X_{1:t} \sim q(X_{1:t}|Y_{1:t}) = q(X_t|Y_t, X_{1:t-1})q(X_{1:t-1}|Y_{1:t})$. The two factors are the proposal probabilities for the new state value, X_t and the history, $X_{1:t-1}$, respectively. Particles are then weighted to take account of the difference between the targeted posterior distribution and the importance distribution:

$$W_t^{(m)} = \frac{P(X_{1:t}^{(m)}|Y_{1:t})}{q(X_{1:t}^{(m)}|Y_{1:t})} \tag{2.26}$$

The simplest choice for the history proposal is $q(X_{1:t-1}|Y_{1:t}) = \mu(X_{1:t-1}|Y_{1:t-1})$, which can be implemented by simply keeping the same set of paths as the previous particle set. This is equivalent to an ordinary IS step with no resampling, and importance weights are given by:

$$W_t^{(m)} = \frac{P(X_{1:t}^{(m)}|Y_{1:t})}{q(X_{1:t}^{(m)}|Y_{1:t})} = \frac{P(X_{1:t}^{(m)}|Y_{1:t})}{\mu(X_{1:t-1}^{(m)}|Y_{1:t-1})q(X_t^{(m)}|X_{t-1}^{(m)}, Y_t)} \approx \frac{W_{t-1}^{(m)} P(Y_t|X_t^{(m)})P(X_t^{(m)}|X_{t-1}^{(m)})}{q(X_t^{(m)}|X_{t-1}^{(m)}, Y_t)} \tag{2.27}$$

Alternatively, we could use $q(X_{1:t-1}|Y_{1:t}) = \hat{P}(X_{1:t-1}|Y_{1:t-1})$ as the history proposal, i.e. sample from the weighted particle distribution which approximates the previous posterior. This is equivalent to an IS step preceeded by resampling. Importance weights are now given by:

$$W_t^{(m)} = \frac{P(X_{1:t}^{(m)}|Y_{1:t})}{q(X_{1:t}^{(m)}|Y_{1:t})} \approx \frac{P(X_{1:t}^{(m)}|Y_{1:t})}{P(X_{1:t-1}^{(m)}|Y_{1:t-1})q(X_t^{(m)}|X_{t-1}^{(m)}, Y_t)} \approx \frac{P(Y_t|X_t^{(m)})P(X_t^{(m)}|X_{t-1}^{(m)})}{q(X_t^{(m)}|X_{t-1}^{(m)}, Y_t)} \tag{2.28}$$

2.4.1 Auxiliary sampling

We can generalise the form of our history proposal distribution by weighting the particles from the previous posterior distribution with any arbitrary set of weights.

$$q(X_{1:t-1}|Y_{1:t}) = \frac{1}{N}\sum_m V_t^{(m)} \delta_X(x_{1:t}^{(m)}) \tag{2.29}$$

Now we have

$$\hat{P}(X_{1:t-1}^{(m)}|Y_{1:t-1}) = \frac{W_{t-1}^{(m)}}{V_t^{(m)}} q(X_{1:t-1}^{(m)}|Y_{1:t}) \tag{2.30}$$

giving a general form for the importance weights

$$W_t^{(m)} = \frac{P(X_{1:t}^{(m)}|Y_{1:t})}{q(X_{1:t}^{(m)}|Y_{1:t})} = \frac{P(X_{1:t}^{(m)}|Y_{1:t})}{\mu(X_{1:t-1}^{(m)}|Y_{1:t-1})q(X_t^{(m)}|X_{t-1}^{(m)}, Y_t)} \approx \frac{W_{t-1}^{(m)}}{V_t^{(m)}} \times \frac{P(Y_t|X_t^{(m)})P(X_t^{(m)}|X_{t-1}^{(m)})}{q(X_t^{(m)}|X_{t-1}^{(m)}, Y_t)} \tag{2.31}$$

2.4.2 Degeneracy and resampling

REWRITE THIS ENTIRELY

Which of the two choices of history proposal should we use? The first, $\mu(X_{1:t-1}|Y_{1:t-1})$, is the simplest to implement, because we can simply keep the same set of paths for $1 : t-1$. However, the recursive form of the importance weights means that the variance of these weights

will increase over time. The result is the well-documented particle degeneracy effect, whereby all the weight coallesces in one or a few particles, and the weights of the rest tend to zero. Intuitively, this is a poor representation of the distribution - we may as well not have the zero-weight particles! The solution is to use resampling, or in the formulation posed above to use the second form of history proposal, $\hat{P}(X_{1:t-1}|Y_{1:t-1})$. This biases the sampling of histories towards those with high weights, encouraging those with low weights to be discarded. Because the importance weights for this type of proposal are not calculated recursively, the variance is reduced.

Particle degeneracy can be quantified ESS

Resampling strategies - systematic, multinomial, etc.

2.4.3 *Importance distributions*

It remains to choose the the importance distribution for the current state, $q(X_t|X_{t-1}, Y_t)$. In the original "bootstrap filter" of (Gordon, Salmond & Smith 1993), this was set equal to the transition density, $P(X_t|X_{t-1})$, which leads to cancellation in the expression for importance weights. This is simple but not necessarily optimal. For example if the process noise is high, the samples of X_t will be widely spread. If, however, the observation noise is comparitively low, many or most of the samples will be far from the observation and will have a low weight. This is undesirable, as discussed in section **??**.

An improved choice of - Optimal importance dist

Add a discussion of the problem of dimensionality.

2.5 MCMC in da house - series vs. parallel

S-MCMC (Godsill) approach

3

Approaches to Tracking

3.1 Why do we track?

In a target tracking scenario, the aim is to trace the trajectory of an object over time from a set of discrete observations. In the previous chapter we saw how we can carry out inference tasks using the recursions of the Kalman filter or the particle filter. Surely now we are done!? What else remains? The additional difficulty in target tracking is the nature of the observation process. Our imperfect sensors may not detect every target present in the scene in every scan. Furthermore, there may be some number of false alarms arising from sensor errors or clutter. Our task thus becomes three-fold: firstly to detect what targets are present in the scene, secondly to work out which observations were generated by each target, and finally to estimate the states of the targets. The basic Kalman and particle filters address only the third of these tasks.

Research in target tracking emerged from military applications such as radar and sonar. However, similar problems emerge in many diverse branches of science: the tracking of people, vehicles or animals in video sequences; of molecules or cells in microscopy data; or of notes in a piece of music. Each can be reduced to a similar underlying model. In this chapter we will formulate such a basic model, keeping it as general as possible rather than focusing on any specific application.

3.2 Models for tracking

In this section we introduce the basic models and notation for the object tracking problem which will be used throughout this report.

For our first assumptions, we will consider our targets to exist at singular points in space, and that targets move independently of each other. Each target will have be characterised by a state vector x_t. This state is composed of continuous-valued coordinates and evolves according to a hidden Markov model (HMM).

$$x_t = f(x_{t-1}, v_t) \tag{3.1}$$

where v_t is a random vector.

The targets are observed by a sensor which detects each with a probability P_D. If detected, the sensor returns a point observation at a location y_t given by:

$$y_t = g(x_t, w_t) \tag{3.2}$$

where w_t is another random vector, independent of v_t.

In addition to the target-originating observations, the sensor also detects a number of false alarms. Throughout this report we assume that these are generated by a Poisson process, with uniform intensity over the observation area.

We denote the set of targets present as $X_t = \{x_{1,t}, x_{2,t}, ..., x_{K_t,t}\}$, and the set of observations as $Y_t = \{y_t^{(1)}, y_t^{(2)}, ..., y_t^{(M_t)}\}$.

3.2.1 Data association

In order to conduct inference on our multi-target system, we would like to calculate the likelihood, $P(Y_t|X_t)$. In order to evaluate this term, we need to hypothesise some assignment between the observations and the process which generated them, whether a particular target or clutter. We introduce an association variable for each target, $\lambda_{j,t}$, which indicates which of the observations in that frame was generated by this target. If the target is not detected, then λ_t is set to 0. We denote the set $\Lambda_t = \{\lambda_{1,t}, \lambda_{2,t}, ..., \lambda_{K_t,t}\}$. As each observations is generated by one target or clutter, no two elements of Λ_t may take the same value, unless 0.

We can now expand the likelihood as:

$$P(Y_t|X_t) = \sum_{\Lambda_t} \prod_{j=1}^{K_t} P(y_t^{(\lambda_{j,t})}|x_{j,t}) \tag{3.3}$$

where the summation is over all feasible values of Λ_t. The number of terms in this summation is $M_t(M_t - 1)...(M_t - K_t)$, which may be prohibitively large if there are many targets or observations in the scene.

If it is necessary to calculate the likelihood very often, as in a particle filter, it may be preferable to estimate Λ_t instead of marginalising it.

3.3 Algorithms for tracking

JPDAF MHT MCMCDA Particle filters. Especially Vermaak RBPF

4

Fixed Lag Estimation for Tracking

4.1 Why do now what we can do later?

In chapter 2, we looked at filtering methods, that is, estimation of a hidden state X_t given observations up until this time $Y_{1:t}$. For such a task, Kalman filters and particle filters are our tools of choice. However, for difficult target tracking problems, there may simply not be enough information available to make a good estimate. Consider a target with a low detection probability and high clutter. There is likely to be multiple observations with which the target could be associated in each frame, or the target could disappear entirely for a period. The result is that there is a very large number of possible routes through the observation space. A conventional particle filter will have to maintain particles on each of the possible routes until such time as they become negligibly unlikely. The correct route may at first only have a few particles on it, and when it is identified as the correct route in later frames the particle diversity may be poor. In the worst case, the correct route may at first look so unlikely (e.g. consecutive missed detections), that no particles follow it, and the track is lost.

This problem can be addressed by allowing not only a proposal of a new state X_t on the end of each path, but also the re-proposal of the preceeding states as well. Thus, particles which have followed an incorret route may be "redirected", to follow a more probable course. This helps maintain particle diversity. If few or no particles follow the correct route at first, particles may be diverted later once the course becomes more apparent.

In an ordinary particle filter, we estimate the joint posterior, $P(X_{1:t}|Y_{1:t})$ at each time instant. Thus, our estimates of the state distribution at previous times are updated at each step, but only by resampling. Diversity in the marginal distributions for previous states only decreases as time progresses. Conversely, in using such a system of re-proposals we allow diversity to be maintained in previous states. To limit the computational complexity, we constrain the region in which re-proposals may occur to a lagging window of L time steps.

4.2 A mathematical framework for fixed lag estimation

We now consider a mathematical framework for fixed lag estimation. This method was devised in (Doucet, Briers & Sénécal 2006) and (Briers, Doucet, Maskell & Horridge 2006).

As before, the target posterior distribution in which we are interested is the familiar $P(X_{1:t}|Y_{1:t})$. However, the proposal mechanism now becomes more complex, because we will be replacing states in an existing particle. We first propose a particle from which to take the state "history", that is $X_{1:t-L}$. This may be chosen from the particle approximation from any of the previous L processing steps. However, we get more of the path than we need, because each particle at lag s is a set of states $X_{1:t-s}$. The final $L-s$ states will be replaced when by a new "tip", $X'_{t-L+1:t}$, drawn from an importance distribution. The complete proposal is thus

$$\{X_{1:t-L}, X'_{t-L+1:t}\} \sim q(s) \int q(X_{1:t-s}|Y_{1:t-s}) q(X'_{t-L+1}|X_{1:t-s}, Y_{t-L+1:t}) dX_{t-L+1:t-s} \tag{4.1}$$

where $q(X_{1:t-s}|Y_{1:t-s})$ is a proposal distribution using the arbitrarily-weighted particles from $\hat{P}(X_{1:t-s}|Y_{1:t-s})$. We cannot evaluate this. In general, the integral will be intractable. If we restrict our proposals to depend only on the history, i.e. use $q(X_{t-L+1}|X_{1:t-L}, Y_{t-L+1:t})$, then the proposal becomes

$$\{X_{1:t-L}, X'_{t-L+1:t}\} \sim q(s) \hat{P}(X_{1:t-L}|Y_{1:t-s}) q(X'_{t-L+1}|X_{1:t-s}, Y_{t-L+1:t}) \tag{4.2}$$

However, with this proposal, in order to evaluate importance weights or acceptance probabilities we still need to calculate

$$P(X_{1:t-L}|Y_{1:t-s}) = \frac{P(Y_{t-L+1:t-s}|X^{(m)}_{t-L}) P(X_{1:t-L}|Y_{1:t-L}}{P(Y_{t-L+1:t-s}|Y_{1:t-L})} \tag{4.3}$$

The problematic term is:

$$P(Y_{t-L+1:t}|X^{(m)}_{t-L}) = \int P(Y_{t-L+1:t}|X^{(m)}_{t-L}, X_{t-L+1:t}) P(X_{t-L+1:t}|X^{(m)}_{t-L}) dX_{t-L+1:t-s} \tag{4.4}$$

In general, this will be intractable. The exception is the case where $s = L$, but this will in general not work well - such a requirement implies ignoring all the estimation we have done since the start of the window!

The solution to this tractability problem proposed in (Doucet et al. 2006) is to augment the dimension of the target distribution to include the discarded tracks. The new target is:

$$P(X_{1:t-L}, X'_{t-L+1:t}|Y_{1:t}) \rho(X_{t-L+1:t-s}|X_{1:t-L}, X'_{t-L+1:t}) \tag{4.5}$$

Once a particle approximation has been generated for this target the required posterior distribution may be obtained by marginalisation. As we are simply going to discard the old track sections $X_{t-L+1:t-s}$, the choice of $\rho(.)$ does not alter the distribution of the posterior. However, the variance of weights/acceptance probabilities may be affected.

With an expanded space for the target distribution there is no need to marginalise the previous states. The proposal distribution now becomes:

$$\{X_{1:t-s}, X'_{t-L+1:t}\} \sim q(s) q(X_{1:t-s}|Y_{1:t-s}) q(X'_{t-L+1}|X_{1:t-s}, Y_{t-L+1:t}) \tag{4.6}$$

4.2.1 Artificial conditionals

We use $\rho(.) = q(.)$. Why?

4.2.2 History Proposals

"Conservative" resampling as a history proposal.

4.3 Applying the fixed lag method to the tracking model

Erm... i've kinda said what I was going to say here in the previous section

4.3.1 Importance distributions

The dimensionality of a fixed lag particle filter will generally be very large. Consider a problem where each target has a 4-dimensional kinematic state (postion and velocity in x and y) and an observation-association index. This gives us five dimensions per target per time step. If we use a lag window with length $L = 5$ and 5 targets we will have a 125 dimensional state. If we used a basic, bootstrap approach to such a problem, the chances of even a single particle following the correct path are negligible. Instead we must exploit the strong correlations between states arising from the structure of the problem. We first factorise the proposal thus:

$$\begin{aligned} q(x_{t-L+1:t}, \lambda_{t-L+1:t} | x_{t-L}, Y_{t-L+1:t}) & \\ &= q(\lambda_{t-L+1:t} | x_{t-L}, Y_{t-L+1:t}) \\ &= q(x_{t-L+1:t} | x_{t-L}, \lambda_{t-L+1:t}, Y_{t-L+1:t}) \end{aligned} \tag{4.7}$$

Thus we can sample first the association variables then the state variables. Each of these terms can be further factorised over time, as we shall see. If we were to replace each of the factors with its corresponding posterior distribution then this factorsiation would recreate the "optimal" importance distribution.

Association proposals

The proposal for the association variables is a discrete distribution over the possible observations with which the target could be associated. Within the factorisation above, the "optimal" form of the proposal is:

$$\begin{aligned} q(\lambda_{t-L+1:t} | x_{t-L}, Y_{t-L+1:t}) & \\ &\propto P(Y_{t-L+1:t} | \lambda_{t-L+1:t}, x_{t-L}) P(\lambda_{t-L+1:t} | x_{t-L}) \\ &= \prod_{\substack{k=1:L \\ tt=t-L+k}} P(Y_{tt} | Y_{tt+1:t} \lambda_{tt:t}, x_{t-L}) P(\lambda_{tt}) \\ &= \prod_{\substack{k=1:L \\ tt=t-L+k}} \int P(Y_{tt} | x_{tt}, \lambda_k) P(x_{tt} | Y_{tt+1:t}, \lambda_{tt+1:t}, x_{t-L}) dx_{tt} P(\lambda_{tt}) \end{aligned} \tag{4.8}$$

This suggests a convenient sequential sampling procedure, starting with λ_t and working backwards in time. The nomalisation constant is not known, but as this is discrete distribution we can enforce normalisation by dividing by the sum.

First we consider a factor from this expression with $\lambda_t t = 1$, i.e. a proposal that in a particular frame the target is not detected. In this case, the observation density is independent of the state of the target, and we have:

$$P(Y_{tt}|Y_{tt+1:t}\lambda_{tt:t}, x_{t-L})P(\lambda_{tt}) = V^{-1}P(\lambda_t = 0) \tag{4.9}$$

When the target is detected, the factors of the proposal ditribution of equation 4.8 may calculated analytically for the linear-Gaussian case. For other cases, the EKF approximations may be used. As this is only a proposal distribution, such an approximation will not affect the distribution of the particle distribution generated. The state distribution term over x_{tt} is problematic, requiring k integrals to calculate. Although this is will be analytic with Gaussian dynamics, its complexity may become unmanageable. This term represents the probability of the state given the set of observations associated with this target at later times. We can render this calculation more manageable by replacing the whole set of future observations with just one.

$$P(x_k|Y_{k+1:t}, \lambda_{k+1:t}, x_{t-L}) \approx P(x_k|Y_{k+d}, \lambda_{k+d}, x_{t-L}) \tag{4.10}$$

For cases with low observation noise, where an observation gives us significant information about the state of the target, this substitution will have little effect. Later observations cannot add much additional information. Again, as this is a proposal distribution, such a substitution will not affect the validity of the resulting particle distribution.

In general we will use $d = 1$, as the closest observation in time will give us the most information about x_{tt}. However, if the target is not detected at time $tt+1$, then we can increase d to find the next detection of the target.

Using this approximation, for $\lambda_{t-L+k} \neq 0$, we have

$$P(Y_{tt}|Y_{tt+1:t}\lambda_{tt:t}, x_{t-L})P(\lambda_{tt}) \propto \mathcal{N}(y_{tt}^{(\lambda_{tt})}|m_{tt}, S_{tt}) \tag{4.11}$$

where usually

$$S_{tt} = [I - R^{-1}C_{tt}\Sigma_{tt}C_{tt}R^{-1}]^{-1} \tag{4.12}$$

$$m_{tt} = SR^{-1}C_{tt}\Sigma_{tt}[(A^d)^T C_{tt+d}^T R_d^{-1} y_{tt+d}^{\lambda_{tt+d}} + Q_k^{-1}A^k x_{t-L}] \tag{4.13}$$

$$\Sigma_{tt} = [C_{tt}^T R^{-1} C_{tt} + (A^d)^T C_{tt+d}^T R_d^{-1} C_{tt+d} A^d + Q_k^{-1}]^{-1} \tag{4.14}$$

$$Q_d = \sum_{l=0}^{d-1} A^l Q (A^l)^T \tag{4.15}$$

$$R_d = R + C_{tt+d} Q_d (C_{tt+d})^T \tag{4.16}$$

We will need a different expression for the case when $tt = t$, because no future associations have yet been proposed. Similarly, if $tt < t$ but the future associations have all been proposed as missed detections, then there are no future observations to guide us, whatever choice of d we use. In these cases we have:

$$S_{tt} = R_d \tag{4.17}$$

$$m_{tt} = C_{tt} A^k x_{t-L} \tag{4.18}$$

For full derivations, see Appendix.
Finally, substituting for the association prior terms, we have:

$$q(\lambda_{t-L+1:t}|x_{t-L}, Y_{t-L+1:t}) \propto \prod_{\substack{k=1:L \\ tt=t-L+k}} \begin{cases} P_D \mathcal{N}(y_{tt}^{(\lambda_{tt})}|m_{tt}, S_{tt}) & \lambda_{tt} = 0 \\ (1-P_D)\mu_C V^{-1} & \lambda_{tt} \neq 0 \end{cases} \tag{4.19}$$

This gives us a complete sequential mechanism for proposing the asociations.

State proposals

Once the associations are fixed, the states can be proposed. When the state space model is linear-Gaussian, we can propose directly from the "optimal" importance distribution for the states using the forward-filtering-backward-sampling algorithm of (Chib 1996), as suggested in (Doucet et al. 2006). For nonlinear models, we can use EKF approximations, as for the associations. Once again, we factorise the proposal:

$$\begin{aligned} &q(x_{t-L+1:t}|x_{t-L}, \lambda_{t-L+1:t}, Y_{t-L+1:t}) \\ &\qquad = P(x_{t-L+1:t}|x_{t-L}, \lambda_{t-L+1:t}, Y_{t-L+1:t}) \\ &\quad = P(x_t|\lambda_{t-L+1:t}, Y_{t-L+1:t}, x_{t-L}) \prod_{k=t-L+1}^{t-1} P(x_k|\lambda_{t-L+1:k}, Y_{t-L+1:k}, x_{t-L}, x_{k+1}) \end{aligned} \tag{4.20}$$

where

$$P(x_k|\lambda_{t-L+1:k}, Y_{t-L+1:k}, x_{t-L}, x_{k+1}) \propto P(x_{k+1}|x_k) P(x_k|\lambda_{t-L+1:k}, Y_{t-L+1:k}, x_{t-L}) \tag{4.21}$$

The distributions $P(x_k|\lambda_{t-L+1:k}, Y_{t-L+1:k}, x_{t-L})$ are given by a Kalman filter, and are Gaussian with mean μ_k and covariance Σ_k. Thus the complete state proposal is given by:

$$q(x_{t-L+1:t}|x_{t-L}, \lambda_{t-L+1:t}, Y_{t-L+1:t}) = \mathcal{N}(x_t|\mu_t, \Sigma_t) \prod_{k=t-L+1}^{t-1} \mathcal{N}(x_k|m_k, S_k) \tag{4.22}$$

where

$$S_k = [A^T Q^{-1} A + \Sigma^{-1}]^{-1} \tag{4.23}$$

$$m_k = S_k[A^T Q^{-1} x_{k+1} + \Sigma^{-1} \mu_k] \tag{4.24}$$

5

Fixed Lag Particle Filters

6

Plan Of Future Research

radar, insects, crowds, cells, jump-diffusions, cursor tracking...

Bibliography

Bayes, M. & Price, M. (1763). An essay towards solving a problem in the doctrine of chances. by the late rev. mr. bayes, frs communicated by mr. price, in a letter to john canton, amfrs, *Philosophical Transactions* **53**: 370.

Briers, M., Doucet, A., Maskell, S. & Horridge, P. (2006). Fixed-lag sequential monte carlo data association, *Proceedings of SPIE*, Vol. 6236, p. 62360S. Paper copy from British Library.

Cappé, O., Godsill, S. & Moulines, E. (2007). An overview of existing methods and recent advances in sequential monte carlo, *Proceedings of the IEEE* **95**(5): 899–924.

Chib, S. (1996). Calculating posterior distributions and modal estimates in markov mixture models, *Journal of Econometrics* **75**(1): 79 – 97. Details forward-filtering-backward-sampling algorithm.

Daum, F. (2005). Nonlinear filters: beyond the kalman filter, *Aerospace and Electronic Systems Magazine, IEEE* **20**(8): 57 –69.

Doucet, A., Briers, M. & Sénécal, S. (2006). Efficient block sampling strategies for sequential monte carlo, *Journal of Computational and Graphical Statistics* **15**(3): 693–711.

Doucet, A. & Johansen, A. M. (2009). A tutorial on particle filtering and smoothing: Fifteen years later, *in* D. Crisan & B. Rozovsky (eds), *The Oxford Handbook of Nonlinear Filtering*, Oxford University Press.

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

Kalman, R. E. (1960). A new approach to linear filtering and prediction problems, *Journal Of Basic Engineering* **82**(Series D): 35–45.

Pitt, M. K. & Shephard, N. (1999). Filtering via simulation: Auxiliary particle filters, *Journal of the American Statistical Association* **94**(446): pp. 590–599.

Rauch, H., Tung, F. & Striebel, C. (1965). Maximum likelihood estimates of linear dynamic systems, *AIAA journal* **3**(8): 1445–1450.