Multi-Target Out-of-Sequence Data Association

Simon Maskell, Richard Everitt, Robert Wright, Mark Briers
QinetiQ
St. Andrews Road
Malvern WR14 3PS, Worcestershire
UK
{s.maskell,r.everitt,r.wright,m.briers}@signal.qinetiq.com

Abstract – *In data fusion systems, one often encounters measurements of past target locations and then wishes to deduce where the targets are currently located. Recent research on the processing of such* out-of-sequence *data has culminated in the development of a number of algorithms for solving the associated tracking problem. This paper reviews these different approaches in a common Bayesian framework and proposes an architecture that orthogonalises the data association and out-of-sequence problems such that any combination of solutions to these two problems can be used together. The emphasis is not on advocating one approach over another on the basis of computational expense, but rather on understanding the relationships between the algorithms so that any approximations made are explicit.*

Keywords: Sensor fusion, tracking, data association, out-of-sequence measurements, particle filter, Kalman filter.

1 Introduction

There are a number of data fusion scenarios which result in a multi-target tracker receiving measurements of where the targets were previously located. In the general case, this is due to the identity of the sensor that generates the measurements of a target affecting the length of time for the information of where the target was to propagate to the tracker. This can be caused by a number of specific effects: communications delays from the sensor to the tracker could be sensor dependent; acoustic propagation could result in different sensors observing the current state of the target at different times; formation of tracks at a sensor could introduce delay in sending these tracks to the fusion node (often because the sensor is a rotating radar with measurement-specific time stamps).

The result is that tracking in the presence of out-of-sequence measurements has received considerable interest over recent years and a number of different approaches have been proposed[1–4]. Here, these approaches are reviewed in the context of a common Bayesian definition of the problem that is similar to that described in [3, 4]. The common definition gives insight into the relative merits of previous approaches and makes it possible to extend these approaches to provide efficient solutions to the general multi-target out-of-sequence data-association problem. These solutions are efficient in that they exploit the structure of the underlying (Markov-chain) model for the targets. Much of the research is motivated by the desire to be able to process out-of-sequence data more efficiently than an algorithm that simply reprocesses all the data. It should be acknowledged that several of the algorithms discussed can be criticised as not achieving this aim. However, the pursuit of a common understanding and the use of only explicit approximation motivates this paper.

So, in section 2 a framework is described that is capable of describing the previous approaches to solving the problems associated with out-of-sequence measurements. Section 3 demonstrates that this framework can accommodate the wide range of existing algorithms and discusses these algorithms within this framework. It is shown, in sections 4 and 5 respectively, to be straightforward to extend this framework to consider data association and multiple targets. Finally, conclusions are drawn in section 6.

2 Stochastic Dynamic Systems

To begin, consider a single target. At time, τ, the state of the target is x_τ. There exists an equation, which enables the future state of the target, $x_{\tau'}$, to be described in terms of the previous state and some random quantity, $\epsilon_{\tau'-\tau}$:

$$x_{\tau'} = f\left(x_\tau, \epsilon_{\tau'-\tau}\right) \tag{1}$$

At one of a number of iterations, k, a measurement, y_k, is received. This measurement is described in terms of the state at the corresponding time, x_{τ_k}, and some random quantity, ω_k:

$$y_k = h\left(x_{\tau_k}, \omega_k\right) \tag{2}$$

Given these equations, one can infer the associated probability densities, namely $p\left(x_{\tau'}|x_\tau\right)$ and $p\left(y_k|x_{\tau_k}\right)$; a minor point is that equations 1 and 2 may convey more information than the probability densities though for all estimation problems, including all tracking problems, the two descriptions can be considered equivalent[1]. These densities define the probability density over the joint distribution of the trajectory of the target and the measurements received up to

[1]Equations 1 and 2 define the physical causal process rather than just the resulting dependencies.

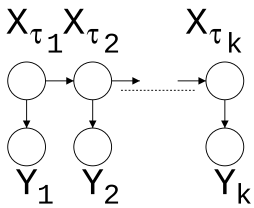

Fig. 1: A stochastic dynamic system for $p\left(x_{\tau_1:\tau_K}, y_{1:K}\right)$.

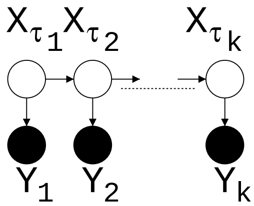

Fig. 2: A stochastic dynamic system for $p\left(x_{\tau_1:\tau_K} | y_{1:K}\right)$.

some point, K. This density can be factorised as follows:

$$p\left(x_{\tau_1:\tau_K}, y_{1:K}\right) = p\left(x_{\tau_1:\tau_K}\right) p\left(y_{1:K} | x_{\tau_1:\tau_K}\right) \tag{3}$$

$$= p\left(x_{\tau_1}\right) \prod_{k=2}^{K} p\left(x_{\tau_k} | x_{\tau_{k-1}}\right) \prod_{k=1}^{K} p\left(y_k | x_{\tau_k}\right) \tag{4}$$

where $p\left(x_{\tau_1}\right)$ is some initial prior, $y_{1:K} = \{y_1, y_2, \ldots y_K\}$ is the history of (as yet unobserved) measurements and $x_{\tau_1:\tau_K} = \{x_{\tau_1}, x_{\tau_2}, \ldots, x_{\tau_K}\}$ is the history of the states at the times corresponding to these measurements. This *stochastic dynamic system* can be drawn diagrammatically as graphical model, a set of interconnected nodes as in figure 1. The nodes represent quantities of interest and the arrows indicate dependencies of one variable on another. The reader unfamiliar with graphical models should refer to one of the many texts available on the subject[5]. Some basic understanding of such models will be assumed from this point.

The graphical model identifies the dependency structure of the variables. The arrows point from child nodes to parent nodes and convey the dependency structure. This specific kind of Markov chain structure is such that the current state is a sufficient statistic of the past; if one knows the current state exactly, there is no more information to be gained by knowing previous states of the system.

Often one observes the true value of a variable and then one wants to update the distribution over all the other unobserved variables (which are referred to as *hidden* in the graphical model literature) in the light of this observation. With the model we are considering, we could imagine that the measurements, $y_{1:K}$, could all be received at once and one would then desire the distribution of the other unobserved variables (the states) given the measurements, ie. $p\left(x_{\tau_1:\tau_K} | y_{1:K}\right)$. In the diagrammatic language of graphical models, this is represented using filled nodes as shown in figure 2.

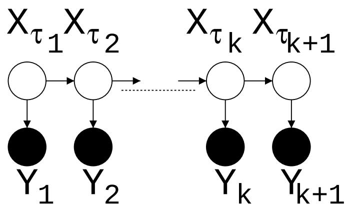

Fig. 3: A stochastic dynamic system for $p\left(x_{\tau_1:\tau_{K+1}} | y_{1:K+1}\right)$.

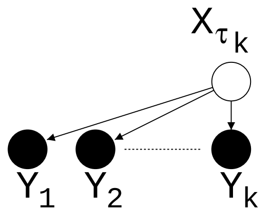

Fig. 4: A stochastic dynamic system for $p\left(x_{\tau_K} | y_{1:K}\right)$.

The dependency structure means that $p\left(x_{\tau_1:\tau_K} | y_{1:K}\right)$ can be expressed very succinctly; one does not need to consider the joint distribution explicitly, but can fully define the distribution using conditional distributions of parent (state) nodes on child (state) nodes. This structure is that exploited by fixed-lag smoothing algorithms such as the Kalman[6] and particle smoother[7]. Note that such joint distributions are often highly correlated; uncertainty over the value of a given state is strongly affected by the values of its neighbours (in time).

When the measurements are received in time order, the chain can simply be augmented with the new nodes for the state and the measurement. The fact that the new measurement is then observed could then be used to deduce the parameters of the chain. The new chain that would result is shown in figure 3.

Removing an unobserved node from the network is equivalent to integrating out the corresponding variable from the joint distribution. When this is conducted, the children of the node become new children of the parents of the node; the parent node gets its old grandchildren as its new children! In tracking, one is only interested in the filtered distribution, $p\left(x_{\tau_K} | y_{1:K}\right)$. So, one can integrate out the nodes corresponding to each of the other states of the target in turn. The resulting chain is shown in figure 4.

One can then augment this chain to form $p\left(x_{\tau_K}, x_{\tau_{K+1}}, y_{K+1} | y_{1:K}\right)$ and then $p\left(x_{\tau_K}, x_{\tau_{K+1}} | y_{1:K+1}\right)$ and $p\left(x_{\tau_{K+1}} | y_{1:K+1}\right)$, which are shown in figures 5, 6 and 7 respectively.

This process is no more than that conducted by a Kalman filter or particle filter, which both store $p\left(x_{\tau_K} | y_{1:K}\right)$ as a sufficient statistic of $p\left(x_{\tau_1:\tau_K} | y_{1:K}\right)$ in terms of the capacity to calculate $p\left(x_{\tau_{K+1}} | y_{1:K+1}\right)$.

2.1 Out-of-Sequence Measurements

The presence of out-of-sequence measurements complicates this process. An out-of-sequence measurement refers

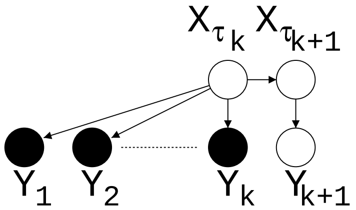

Fig. 5: A stochastic dynamic system for $p\left(x_{\tau_K}, x_{\tau_{K+1}}, y_{K+1}|y_{1:K}\right)$.

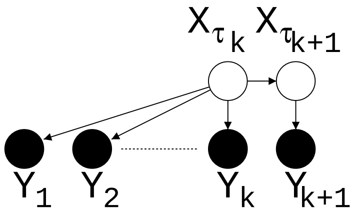

Fig. 6: A stochastic dynamic system for $p\left(x_{\tau_K}, x_{\tau_{K+1}}|y_{1:K+1}\right)$.

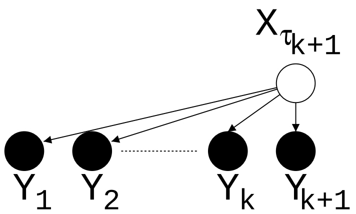

Fig. 7: A stochastic dynamic system for $p\left(x_{\tau_{K+1}}|y_{1:K+1}\right)$.

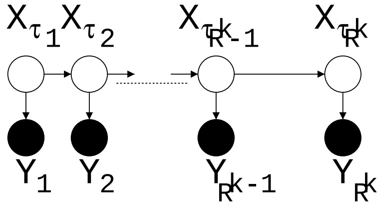

Fig. 8: A stochastic dynamic system for $p\left(x_{\tau_1:\tau_K}|y_{1:K}\right)$.

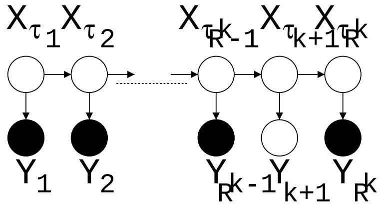

Fig. 9: A stochastic dynamic system for $p\left(x_{\tau_1:\tau_{K+1}}, y_{K+1}|y_{1:K}\right)$.

to a state at a point in the past; if the $(K+1)$th measurement is the first out-of-sequence measurement to be received then $\tau_{K+1} < \tau_K$. In the more general case, the quantity of interest is often then $p\left(x_{\tau_{R^{K+1}}}|y_{1:K+1}\right)$, where $\tau_{R^{K+1}} \geq \tau_{K'}$ for all $K' \leq K+1$ so $\tau_{R^{K+1}}$ is the time of the most *R*ecent in time measurement. Essentially, one wants to revise the belief about where the target is given a measurement of where it was.

In this environment, $p\left(x_{\tau_{R^K}}|y_{1:K}\right)$ is not a sufficient statistic of $p\left(x_{\tau_1:\tau_K}|y_{1:K}\right)$ in terms of the capacity to calculate $p\left(x_{\tau_{R^{K+1}}}|y_{1:K+1}\right)$. However, it is still possible to augment the chain (shown in terms of R^K in figure 8) to form $p\left(x_{\tau_1:\tau_{K+1}}, y_{K+1}|y_{1:K}\right)$ and then $p\left(x_{\tau_1:\tau_{K+1}}|y_{1:K+1}\right)$, which are shown in figures 9 and 10 respectively.

So, this means that on receipt of an out-of-sequence measurement, there is a need to update the parameters of the chain. A naive implementation of an out-of-sequence measurement processing algorithm then necessitates reprocessing of the data from the time of the out-of-sequence measurement to the last time. Research has focused on investigating the potential for more efficient alternative schemes.

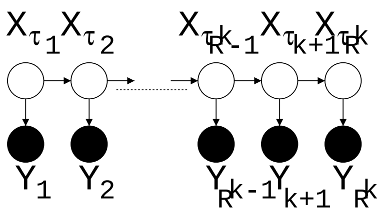

Fig. 10: A stochastic dynamic system for $p\left(x_{\tau_1:\tau_{K+1}}|y_{1:K+1}\right)$.

3 Alternatives to Re-Processing

So, there exist a number of algorithms that attempt to process out-of-sequence measurements efficiently (and at all; a widely advocated approach appears to be to simply ignore such measurements!). These algorithms can be divided into two broad groups: (a) those that store the parameters of the chain and (b) those that do not and so necessitate processing of data that are earlier than the start of the chain. These two groups will be considered in turn. The emphasis is on the models used and not on the details of the matrix manipulations that result under certain assumptions regarding the models and the algorithms then used.

It is worth noting that, in case of linear Gaussian models, it is true to say that as the time of an out-of-sequence measurement becomes increasingly far into the past, the effect on the filtered distribution reduces. However, with non-linear models, this does not necessarily follow; an out-of-sequence measurement could, for example, drastically change the probabilities of different association events and so have a large effect even at a long lag.

3.1 Within Chain Algorithms

One approach to representing the uncertainty over the chain is to consider a *stacked-state*, X_{τ_K}, which consists of the states over a fixed lag, L, stacked on top of one another, $x_{\tau_{R^k-L}:\tau_{R^k}}$. Out-of-sequence measurements that fall into this lag then appear as in-sequence with respect to this stacked-state[4] so one can use standard tracking algorithms to track this stacked-state. However, this approach does not exploit the structure of the model and stores a verbose description of the fixed-lag distribution. In the case considered by [4] of linear Gaussian models, the Kalman filter is the optimal tracking algorithm in terms of its capacity to describe the pdf. However, by storing the full covariance matrix of the stacked-state, more parameters are stored than those required to completely describe the fixed-lag distribution. It should be said that [4] does propose some approximation strategies to improve efficiency, but the authors believe that the fact that the fixed-lag distribution is not efficiently parameterised means that the approach suffers in terms of efficiency.

3.1.1 Algorithms A and B

If the out-of-sequence measurement is between the times corresponding to the last two nodes in the fixed lag distribution, such that $\tau_{R^K-1} \leq \tau_{K+1} \leq \tau_{R^K}$ (so $R^{K+1} = R^K$) then the only parts of the chain that need to be stored to obtain $p\left(x_{\tau_{R^K}}|y_{1:K}\right)$ are those relating to the final two nodes. So, one can augment $p\left(x_{\tau_{R^K-1}}, x_{\tau_{R^K}}|y_{1:K}\right)$ to produce $p\left(x_{\tau_{R^K-1}}, x_{\tau_{K+1}}, x_{\tau_{R^K}}, y_{K+1}|y_{1:K}\right)$, and then $p\left(x_{\tau_{R^{K+1}}}, y_{K+1}|y_{1:K}\right)$ and so $p\left(x_{\tau_{R^{K+1}}}|y_{1:K+1}\right)$, which are respectively shown in figures 11, 12, 13 and 14.

This is what algorithm A implements[2]. Algorithm A assumes linear Gaussian models and that $R^K - 1 = R^{K-1}$ so one can parameterise $p\left(x_{\tau_{R^K-1}}, x_{\tau_{R^K}}|y_{1:K}\right)$ using the last two filtered distributions, $p\left(x_{\tau_{R^K}}|y_{1:K}\right)$ and

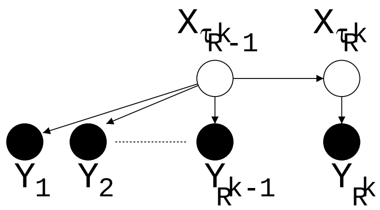

Fig. 11: A stochastic dynamic system for $p\left(x_{\tau_{R^K-1}}, x_{\tau_{R^K}}|y_{1:K}\right)$.

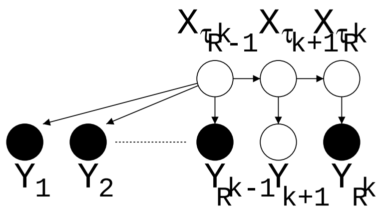

Fig. 12: A stochastic dynamic system for $p\left(x_{\tau_{R^K-1}}, x_{\tau_{K+1}}, x_{\tau_{R^K}}, y_{K+1}|y_{1:K}\right)$.

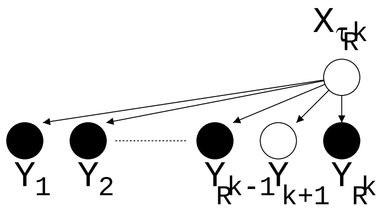

Fig. 13: A stochastic dynamic system for $p\left(x_{\tau_{R^{K+1}}}, y_{K+1}|y_{1:K}\right)$.

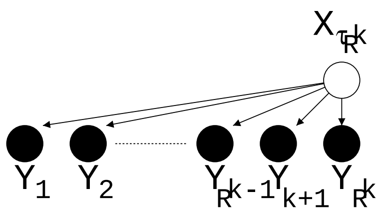

Fig. 14: A stochastic dynamic system for $p\left(x_{\tau_{R^{K+1}}}|y_{1:K+1}\right)$.

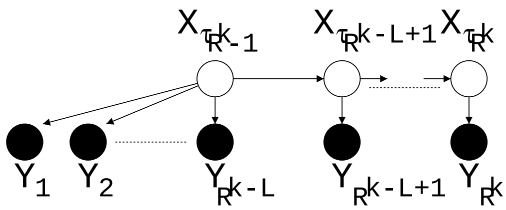

Fig. 15: A stochastic dynamic system for $p\left(x_{\tau_{R^{K}-L}:\tau_{R^{K}}}|y_{1:K}\right)$.

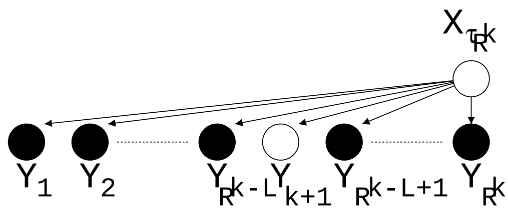

Fig. 17: A stochastic dynamic system for $p\left(x_{\tau_{R^{K+1}}}, y_{K+1}|y_{1:K}\right)$.

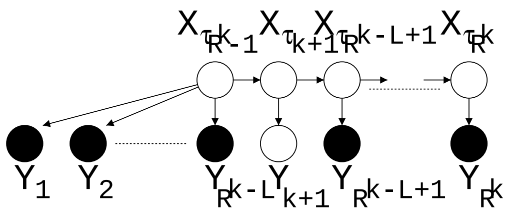

Fig. 16: A stochastic dynamic system for $p\left(x_{\tau_{R^{K}-L}:\tau_{R^{K}}}, x_{\tau_{K+1}}, y_{K+1}|y_{1:K}\right)$.

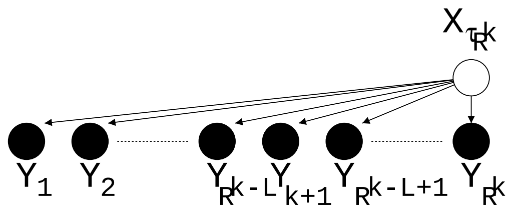

Fig. 18: A stochastic dynamic system for $p\left(x_{\tau_{R^{K+1}}}|y_{1:K+1}\right)$.

$p\left(x_{\tau_{R^{K-1}}}|y_{1:K-1}\right)$, and some other quantities calculated during the filtering process (that account for the conditional dependency structure). If the assumptions hold, algorithm A is exact in its capacity to deduce $p\left(x_{\tau_{R^{K+1}}}|y_{1:K+1}\right)$ from $p\left(x_{\tau_{R^{K}-1}}, x_{\tau_{R^{K}}}|y_{1:K}\right)$.

However, if $R^{K}-1 \neq R^{K-1}$, so the last two filtered distributions don't refer to the last two times, the algorithm will be approximate; the stored filtered distributions will then not be able to exactly describe $p\left(x_{\tau_{R^{K}-1}}, x_{\tau_{R^{K}}}|y_{1:K}\right)$. This situation can arise if two out-of-sequence measurements are received one after the other such that $R^{K+2}=R^{K}$.

Algorithm B[2] then introduces an approximation that results in a minor reduction in computational cost.

3.1.2 Algorithms Al and Bl

In the general case, the out-of-sequence measurement is between the times corresponding to two nodes in the fixed lag distribution. So, this fixed lag distribution then needs to be stored for some lag, L. Here we assume that $\tau_{R^{K}-L} < \tau_{K+1} < \tau_{R^{K}-L+1}$; the lag is just long enough to accommodate the out-of-sequence measurement. If this distribution is stored, one can augment $p\left(x_{\tau_{R^{K}-L}:\tau_{R^{K}}}|y_{1:K}\right)$ to produce $p\left(x_{\tau_{R^{K}-L}:\tau_{R^{K}}}, x_{\tau_{K+1}}, y_{K+1}|y_{1:K}\right)$, and then $p\left(x_{\tau_{R^{K+1}}}, y_{K+1}|y_{1:K}\right)$ and so $p\left(x_{\tau_{R^{K+1}}}|y_{1:K+1}\right)$, which are respectively shown in figures 15, 16, 17 and 18.

In a similar manner to previously, if $R^{K}-L'=R^{K-L'}$ for $L' \leq L$, so the last L filtered distributions refer to the last L times, then $p\left(x_{\tau_{R^{K}-L}:\tau_{R^{K}}}|y_{1:K}\right)$ can be parameterised using stored filtered distributions and other quantities calculated when conducting the filtering. So, it is then possible to deduce $p\left(x_{\tau_{R^{K+1}}}|y_{1:K+1}\right)$ from the stored $p\left(x_{\tau_{R^{K}-L'}}|y_{1:K-L'}\right)$ for all $L' \leq L$. One could then process the out-of-sequence measurement using the chain as parameterised using the stored filtered distributions.

Motivated by the need to reduce storage, one can summarise the effect of the measurements, $y_{R^{K}-L+1:R^{K}}$, on $x_{\tau_{R^{K}}}$ with a single equivalent likelihood function[2], $p\left(y^{\star}|x_{\tau_{R^{K}}}\right)$, which is not technically a pdf since it is parameterised by $x_{\tau_{R^{K}}}$. The use of this summary approximates the fixed lag distribution (shown in figure 15) in such a way that the approaches described in section 3.1.1 can be used to process the out-of-sequence measurement. This approximation is illustrated in figure 19.

The resulting algorithm has been shown to perform well and is known as Algorithm Al[1]. Were the equivalent measurement to parameterise a different likelihood function, namely $p\left(y^{\star}|x_{\tau_{R^{K}-L+1}}, x_{\tau_{R^{K}}}\right)$, the approach could be made exact though the same reduction in storage would then be difficult to achieve. However, since the dependency

[2]The authors use the terminology that a likelihood, $p\left(y|x\right)$, is a pdf parameterised by y with x fixed, while a likelihood function, confusingly often also denoted $p\left(y|x\right)$, is parameterised by x with y fixed.

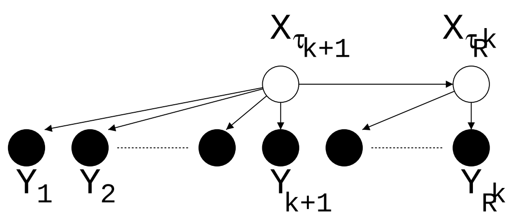

Fig. 19: A stochastic dynamic system for an approximation to $p\left(x_{\tau_{K+1}}, x_{K+1}|y_{1:K+1}\right)$.

structure between the measurements and the state sequence is lost, algorithm Al is approximate. Algorithm Bl employs a further approximation to facilitate a minor reduction in computation[1].

By way of an aside and as observed in [4], it should be noted that this use of an equivalent measurement is the same idea as is used in the context of track fusion to transmit measurement summaries in place of either tracks or measurements. Rather than transmit all the measurements from sensor nodes to a central tracker, individual nodes track the measurements they receive and transmit likelihood functions, which can be thought of as the parameters of a single measurement that would have resulted in the same change to the track as the measurements that were actually received. The track fusion node can then treat these likelihood functions as independent measurements and can therefore employ a tracker to conduct track fusion.

3.1.3 Efficient and Exact Algorithms

The problem that has been hinted at in this discussion is that for these efficient algorithms to be exact, the fixed lag distribution must be completely described. When this is not the case, an efficient optimal update with out-of-sequence measurements is not possible using these approaches; reprocessing of the data appears to be the only efficient and optimal option available if one restricts oneself to algorithms that represent the uncertainty using Gaussian distributions.

An alternative method has been developed using particle filters[3]. Particle filters represent the uncertainty over the history of states using the diversity of a number of hypothesised trajectories through the state space over time. Each hypothesised trajectory has an associated weight. In the context of tracking in the absence of out-of-sequence measurements, the hypothesis of the current state is the sufficient statistic of the history and so this is often the quantity stored by the particle filter. However, the path of the particle through the state space can also be stored; this is not the same as the sequence of filtered distributions. The weights on the particles refer to the relative probabilities of the different hypothesised paths. For each such path, the states at the times of the processed measurements are fixed. So, to process an out-of-sequence measurement, there is no need to reprocess all the subsequent data. The trajectory through the state space is augmented with a sample of the state at the time of the out-of-sequence measurement (which can be a sample from a proposal based on the state at both neighbouring times in the chain) and the weight on the trajectory adjusted.

However, there is a problem. The particle weights represent the disparity between the posterior and the proposal distribution used to sample the trajectory. This disparity necessarily accumulates over time and causes degeneracy. This degeneracy is avoided by using a resampling step, whereby particles are probabilistically replicated and discarded. If you didn't resample, you'd get degeneracy since, as time evolves, the proposal distribution essentially becomes a worse and worse approximation to the fixed-interval posterior distribution. When you resample, you make copies of some of the particles, which means that this fixed-interval posterior isn't well approximated. However, the filtered distribution is well approximated. The result is that the degeneracy that would result is moved to the other end of the trajectory (at times far from the filtered time). So, the resampling operation can be thought of as moving this degeneracy problem down the markov chain. The particles therefore explore a fixed-lag posterior distribution and one can think of a particle filter that uses resampling as an (implicit) fixed-lag smoother.

So, a problem comes about if, as is the case when considering out-of-sequence measurements, the fidelity of the approximation to the fixed-lag distribution is of interest. So, it becomes of increased importance that the proposal distribution is well matched to the fixed-lag posterior. Alternatively, mechanisms for using a sub-optimal proposal distribution can be used. Such mechanisms include the use of rejunivating Metropolis-Hastings steps and less rigorous variants known as jitter and regularisation. However, since the distribution of the chain has a large amount of correlation structure, such moves are necessarily small is size. One could introduce moves on the whole chain, the parameters of which could be deduced through analysis of the correlation structure of the particle cloud, though this would necessitate recalculation of the weights for the fixed-lag and so some form of reprocessing of the measurements. The bottom line is that, when using a particle filter to process out-of-sequence measurements, increased care is needed in the choice of proposal distribution.

3.2 Going Past the Start of the Chain

The discussion to this point has focused on the processing of out-of-sequence measurements that arrive within a fixed-lag of the most recent time. There are situations when one wishes to update the filtered distribution as a result of a measurement that lies outside this lag. Before discussing the algorithms that result, it is necessary to have a brief discussion of reverse-time dynamics.

3.2.1 Reverse-Time Dynamics

The dynamics are often specified in terms of a dynamic model for the forward-time dynamics, that is a function of the form of (1). This function defines a probability distribution, $p\left(x_{\tau'}|x_{\tau}\right)$, over the possible values for $x_{\tau'}$ given the value of x_{τ}.

It is possible to rearrange (1) and so derive an equation for x_{τ} in terms of $x_{\tau'}$:

$$x_{\tau} = f^{-1}\left(x_{\tau'}, \epsilon_{\tau'-\tau}\right) \tag{5}$$

Three issues warrant discussion. First, in the general case, this inverse function doesn't exist or isn't unique. As a specific example of when this happens, consider the case when (1) is of the following form:

$$x_{\tau'} = Fx_{\tau} + \epsilon_{\tau'-\tau} \tag{6}$$

where $\epsilon_{\tau'-\tau}$ is drawn from a zero-mean Gaussian distribution. Then (5) becomes:

$$x_{\tau} = F^{-1}x_{\tau'} + \overleftarrow{\epsilon}_{\tau'-\tau} \tag{7}$$

where if the covariance of $\epsilon_{\tau'-\tau}$ is Q, the covariance of $\overleftarrow{\epsilon}_{\tau'-\tau}$ is $F^{-1}QF^{-1^T}$. If F doesn't have an inverse, then problems will arise. In this specific case, it is possible to avoid the need to calculate F^{-1} and use a parameterisation (using Information Matrices) of the Kalman filter when considering reverse time dynamics (such that only $F^{-1^{-1}} = F$ is needed). However, in the general nonlinear case this isn't possible.

A second issue is that this process of inverting $f(.)$ does not generally result in a probability distribution at all; the process essentially calculates $p(x_{\tau'}|x_\tau)$ as a function of x_τ. This is not guaranteed to integrate to unity over x_τ since the determinant of the transformation could differ from unity and is potentially state dependent. One needs to recall the standard result regarding transformations of random samples from a distribution for z to a distribution for $x = h^{-1}(z)$:

$$p\left(x = h^{-1}(z)\right) = \frac{p(z = h(x))\left|\frac{dh(x)}{dx}\right|}{\int p(z = h(x'))\left|\frac{dh(x')}{dx'}\right| dx'} \tag{8}$$

The implication is that care is needed and that if one simply inverts the functional form of the dynamics, the result is not necessarily a (reverse-time) distribution.

Furthermore, a third issue is that this distribution, even once calculated isn't then $p(x_\tau|x_{\tau'})$! In fact, by Bayes rule:

$$p(x_\tau|x_{\tau'}) = \frac{p(x_{\tau'}|x_\tau)\,p(x_\tau)}{p(x_{\tau'})} \tag{9}$$

where $p(x_\tau)$ and $p(x_{\tau'})$ are the prior on x_τ and $x_{\tau'}$ respectively resulting from the initial prior, $p(x_0)$:

$$p(x_\tau) = \int p(x_0)\,p(x_\tau|x_0)\,dx_0 \tag{10}$$

The use of these priors then ensures that the (forwards and backwards) dynamics are consistent with the initial prior; using the forward-dynamics from the initial prior and then using the backwards-dynamics to go back to the time of the initial prior should leave the initial prior:

$$p(x_0) = \int\left(\int p(x_{0'})\,p(x_\tau|x_{0'})\,dx_{0'}\right)p(x_0|x_\tau)\,dx_\tau \tag{11}$$

$$= \int p(x_\tau)\,p(x_0|x_\tau)\,dx_\tau = p(x_0) \tag{12}$$

3.3 Algorithm C

If one chooses the reverse-time dynamics to be deterministic and so have no associated uncertainty[3] then these issues disappear. This special choice of reverse-time model also means that the fixed-lag distribution does not need to be considered and one can simply store the filtered distribution, augment the chain and then update the filtered distribution. This is one interpretation of Algorithm C[1] and the model for processing of an out-of-sequence measurement is shown in figure 20.

[3] In the case of linear Gaussian models, this is equivalent to approximating $Q = 0$ when calculating the parameters of the reverse-time dynamics.

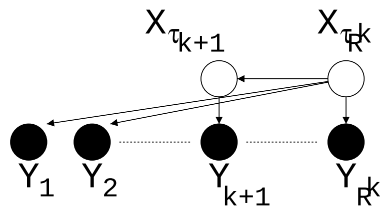

Fig. 20: A stochastic dynamic system for Algorithm C's approximation to $p\left(x_{\tau_{RK+1}}, x_{K+1}|y_{1:K+1}\right)$.

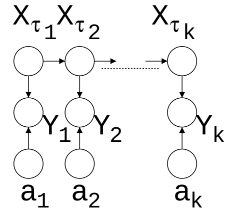

Fig. 21: A stochastic dynamic system for $p(x_{\tau_1:\tau_K}, a_{1:K}, y_{1:K})$.

Evidently, one can devise improved approximation schemes by considering the issues raised in the previous section. A compromise is needed between having a parsimonious representation (so with small storage requirements) and resulting performance in scenarios of interest. These schemes are not discussed here.

4 Data Association

The reason for the completeness of the previous discussion is to illustrate that all the previous approaches to processing out-of-sequence measurements (and some others) can be easily considered as special cases of a generic framework based on explicit modelling of the fixed-lag probability distributions. This section will now show that it is straightforward to describe the data association in the same context and so solve out-of-sequence data association problems.

The association problem can be modelled using another set of nodes in the graphical model. At the Kth iteration, an association variable, a_K, governs the association of the measurements with the target. The graphical model associated with this interpretation of the data association problem is shown in figure 21[4].

Observation of a node, ie. knowledge of the measurements, then results in the graphical model shown in figure 22.

The fact that the measurements are observed means that the distributions for $a_{1:K}$ and $x_{\tau_1:\tau_K}$ are coupled. So, one approach to solving the data association problem is to sample both the state and the association sequence. Another, which exploits the structure of the problem, is to sample one chain and then consider the other chain conditional on this sample. In all such cases each sample has associated

[4] Note that we use circles to represent nodes corresponding to both continuous and discrete variables.

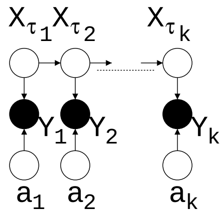

Fig. 22: A stochastic dynamic system for $p\left(x_{\tau_1:\tau_K}, a_{1:K}|y_{1:K}\right)$.

with it a weight and a conditional distribution on the other part of the problem. In the case of a multiple hypothesis tracker, the filter samples (the N most probable) $a_{1:K}$ and then calculates the statistics of $p\left(x_{\tau_K}|a_{1:K}, y_{1:K}\right)$ for each sampled $a_{1:K}$. In the case of a particle filter, it is possible to sample $x_{\tau_1:\tau_K}$ and then calculate the statistics of $p\left(a_{1:K}|x_{\tau_1:\tau_K}, y_{1:K}\right)$ for each sampled $x_{\tau_1:\tau_K}$. Note that, because of the structure of the model, the only flow of dependence from one time step to the next is through the state sequence, $x_{\tau_1:\tau_K}$. Hence, $p\left(a_{1:K}|x_{\tau_1:\tau_K}, y_{1:K}\right)$ factorises through time:

$$p\left(a_{1:K}|x_{\tau_1:\tau_K}, y_{1:K}\right) = \prod_{k=1}^{K} p\left(a_k|x_{\tau_1:\tau_K}, y_{1:K}\right) \qquad (13)$$

One can integrate out a node in the association sequence and so reduce the number of samples in the representation of $p\left(x_{\tau_1:\tau_K}, a_{1:K}|y_{1:K}\right)$. This is the approach taken by the Probabilistic Data Association Filter[8]. Here the interpretation as mixture reduction is made entirely explicit.

So, if one is explicit about sampling the associations or the state and then modelling the resulting effect on the other half of the chain, then the data association and out-of-sequence measurement problems can be orthogonalised in some sense. Just as some structures for modelling out-of-sequence data necessitate what amounts to reprocessing of the data, so one needs to perform the same reprocessing to recalculate the weights on the samples. One could equally explicitly approximate to avoid this reprocessing.

Again, if one uses a particle filter to sample the state sequence, the need to reprocess can be avoided so long as the proposal distribution is chosen judiciously. However, in the presence of uncertain data association, it is difficult to devise such intelligent proposal distributions.

5 Multiple Targets

The various methods for conducting multiple target data association can be posed as special cases of a generic solution strategy based on: calculating a discrete distribution over the candidate values for the association variable for each target; processing these distributions in some way to find the impact of the *best* or the *average* association event for the targets jointly; calculating the effect of some revised discrete distributions on the individual targets[9]. Each target therefore needs to be able to calculate the weights on each of the candidate association events over which the multi-target constraints are to be imposed.

Because of the observation that, by being explicit about the representation of the uncertainty, the data association problem can be orthogonalised from that of processing out-of-sequence data, there is no requirement regarding the time associated with the association events. So, if, as is the case for the work that motivated this review and extension of the literature, one is considering radars that rotate at different rates and report at the end of their scans, one can *ask* each target to produce the distribution over the candidate association events even if these events are stored in a different order internally for one target than for another (or potentially even for one sample for the same target to another such sample).

6 Conclusions

The wide range of algorithms that exist for processing out-of-sequence measurements have been shown to be specific approximations to a generic framework that is described using graphical models. It has shown to be straightforward to extend this framework to cater for out-of-sequence data association with multiple targets.

Acknowledgements

This work was supported by the UK MoD through work funded by the Defence Technology Center for Data and Information Fusion. The authors gratefully acknowledge additional funding from the Royal Commission for the Exhibition of 1851 and thank Csaba Szepesvari for useful discussions regarding transformations of random samples. The authors also wish to thank the reviewers for their helpful comments.

References

[1] Y Bar-Shalom, M Mallick, H Chen, and R Washburn. One-step solution for the general out-of-sequence measurement problem in tracking. In *Proceedings of the 2002 IEEE Aerospace Conference*, March 2002.

[2] Y Bar-Shalom. Update with out-of-sequence measurements in tracking: Exact solution. *IEEE Transactions on Aerospace and Electronic Systems*, 38(3), July 2002.

[3] M Orton and A Marrs. A Bayesian approach to multi-target tracking data fusion with out-of sequence measurements. Technical report, CUED, 2001.

[4] S Challa, R Evans, X Wang, and J Legg. A fixed-lag smoothing solution to out-of-sequence information fusion problems. *Journal of Communications in Information and Systems*, 2(4):325–348, 2002.

[5] R Cowell, P Dawid, S Lauritzen, and D Spiegelhalter. *Probabilistic Networks and Expert Systems*. Springer, 1999.

[6] R H Shumway and D S Stoffer. *Time Series Analysis and Its Applications*. Springer-Verlag: New York, 2000.

[7] A Doucet. On sequential simulation-based methods for Bayesian filtering. Technical Report CUED/F-INFENG/TR 310, Department of Engineering, Cambridge University, 1998.

[8] Y Bar-Shalom and X-R Li. *Multitarget-Multisensor Tracking: Priciples and Techniques*. YBS, 1995.

[9] S Blackman and R Popoli. *Design and Analysis of Modern Tracking Systems*. Artech House, Boston/London, 1-58053-006-0, 1999.