Abstract—This paper proposes a means to achieve tractable particle PHD smoothing through the use of an augmented state space label which tracks the evolution of particles over time. The use of the label reduces the forward-backward particle smoother from quadratic to linear complexity in the number of targets allowing smoothing to be carried out on a large number of targets as well as in the presence of moderate and high levels of clutter.

Keywords: PHD filters, Finite Set Statistics, forward-backward smoothing.

I. INTRODUCTION

The probability hypothesis density (PHD) filter was introduced in [1] as a suboptimal alternative to the recursive Bayesian multitarget filter [2]. The first PHD filter implementations were based on sequential Monte Carlo (SMC) approximations [3]–[5]. Subsequently a closed form solution was discovered by Vo and Ma [6].

More recently, there has been interest in the development of PHD smoothers [7]–[9]. Smoothing [10]–[12] provides a means of improving the quality of the estimates by reducing the localisation error. The first implementation was proposed by Mahler, Vo and Vo [9] using a SMC approximation and subsequently a closed form solution was discovered by Vo and Vo [13].

The SMC implementation described by Mahler, Vo and Vo [9] has a considerable computational complexity due to the requirement of computing multiple integrals over multitarget PHDs. In this paper, we introduce state labelling to identify target trajectories. This enables us to identify the PHD associated with each target and hence, the integrals in the PHD smoother require less computation since we smooth on a track-labelled basis. We show through simulated results that there is a significant reduction in computational complexity, particularly in scenarios with high numbers of targets and high clutter levels.

In the next section, we describe the multitarget Bayes filter and forward-backward smoother and the first moment approximations known as the PHD filter and smoother. In Section III, we describe the SMC implementations of the PHD filter and smoother. In Section IV, we introduce target state labelling. Section V presents results from simulated scenarios and we conclude in Section VI.

II. MULTI-TARGET FILTERING AND SMOOTHING

Multi-target filtering involves estimation of the hidden state of an unknown number of targets given a set of partial observations. Let the hidden target states at time \(k\) be denoted by \(\{x_{k,1}, \ldots, x_{k,m_k}\}\) where \(m_k\) is the number of targets present, each taking a value in the state space \(\mathcal{X} \subseteq \mathbb{R}^{n_x}\). Similarly, let the \(n_k\) observations be given by \(\{z_{k,1}, \ldots, z_{k,n_k}\}\) where each observation exists in \(\mathcal{Z} \subseteq \mathbb{R}^{n_z}\). The multitarget state and observation are expressed as

\[
X_k = \{x_{k,1}, \ldots, x_{k,m_k}\} \in \mathcal{F}(\mathcal{X})
\]

\[
Z_k = \{z_{k,1}, \ldots, z_{k,n_k}\} \in \mathcal{F}(\mathcal{Z})
\]

where \(\mathcal{F}(\mathcal{Y})\) is the class of all finite subsets of \(\mathcal{Y}\).

A random finite set (RFS) \(X\) on \(\mathcal{X}\) is a random variable taking values in \(\mathcal{F}(\mathcal{X})\). This can be used to represent the uncertainty in the state and observation sets. For example, the observation set \(Z_k\) may be composed of measurements due to detected targets as well as clutter (false detections). There is also a possibility that certain targets are not detected and thus these targets do not give rise to an observation.

The optimal multitarget Bayesian filter estimates the posterior distribution conditioned on the observation sets up to time \(k\), \(p_{k|k}(X_k|Z_{1:k})\). Let the multitarget transition density and likelihood function be \(f_{k|k-1}(X_k|X_{k-1})\) and \(g_k(Z_k|X_k)\) respectively. Then, the recursive multitarget Bayesian filter can be expressed as [1]

\[
p_{k|k-1}(X_k|Z_{1:k-1}) = \int p_{k|k-1}(X_k|Z_{1:k-1}) f_{k|k-1}(X_k|X_{k-1}) \delta X_{k-1}
\]

\[
p_{k|k}(X_k|Z_{1:k}) = \frac{g_k(Z_k|X_k)p_{k|k-1}(X_k|Z_{1:k-1})}{\int g_k(Z_k|X)p_{k|k-1}(X|Z_{1:k-1}) \delta X}
\]

where the integrals here represent set integrals [1].

The multitarget forward-backward smoother is given by the expression [8], [9], [14], [15]

\[
p_{k'|k}(X|Z_{1:k}) = \int \left( \frac{f_{k'+1|k'}(Y|X)p_{k'+1|k'}(X)}{\int f_{k'+1|k'}(Y|W)p_{k'+1|k'}(W) \delta W} \right) p_{k'|k}(Y) \delta Y .
\]

where the smoothed posterior density, \(p_{k'|k}(X|Z_{1:k})\), is smoothed backwards from some time \(k\), to \(k' < k\).

The multitarget filter and smoother presented in this form is non-trivial due to the presence of a time-varying number of targets as well as the associated uncertainty in detection and the presence of clutter. The complexity of the recursion grows
exponentially with the number of targets making this approach intractable when large numbers of targets are present.

A. PHD Filtering and Smoothing

The PHD filter [1] was developed as a tractable suboptimal alternative to the multitarget Bayesian filter. The PHD filter provides a first moment approximation of the multitarget posterior distribution on the single target state space. For a RFS on \( \mathcal{X} \) with a probability distribution \( P \), the PHD, or intensity function, is a non-negative function \( v \) on \( \mathcal{X} \) such that the integral of this intensity over any region \( S \subseteq \mathcal{X} \) gives the expected number of elements within that region:

\[
\int |X \cap S| P(dX) = \int_S v(x)dx .
\]

Local maxima in the intensity correspond to regions associated with a higher concentration of elements and these maxima can be used to generate estimates of the elements from \( X \).

Consider a tracking scenario where the single-target state is denoted by \( x \). A probability of detection \( p_{D,k}(x) \) is associated with the target giving rise to a measurement. A sensor detects measurements \( z \in \mathcal{Z}_k \) at time \( k \) and, additionally, the set of observations may consist of false detections due to clutter. If \( D_{k-1|k-1} \) is the updated PHD at time \( k-1 \), then the predicted PHD at time \( k \) is given by the expression [1]

\[
D_{k|k-1}(x) = \gamma_k(x) + \int p_{S,k}(\zeta)D_{k-1|k-1}(\zeta)f_{k|k-1}(x|\zeta)d\zeta (7)
\]

where \( \gamma_k(x) \) is the probability of an existing target surviving to the next time step and \( f_{k|k-1}(x|\zeta) \) is the single target transition density.

The PHD update equation is given by [1]

\[
D_{k|k}(x) = D_{k|k-1}(x)L(Z|x) (8)
\]

where the PHD pseudo-likelihood is given by

\[
L(Z|x) = (1 - p_{D,k}(x))D_{k|k-1}(x) + \sum_{z \in \mathcal{Z}_k} \lambda_kw_k(z|x)D_{k|k-1}(x) (9)
\]

At time \( k \), \( p_{D,k}(x) \) is the probability that a measurement is received from the target with state \( x \), \( g_k(z|x) \) is the single target observation likelihood function, \( c_k(z) \) is the probability distribution of the clutter and \( \lambda_k \) is the average number of clutter points.

The multitarget smoother given by equation (5) encounters the same intractability inherent in the multitarget filter. Smoothing based on the PHD has been proposed as a tractable alternative to achieve multitarget smoothing [7]–[9] which uses the PHD to approximate the forward filter and a backward PHD recursion to approximate the smoothed posterior. The expression for the backward smoothing recursion is given by [8], [9]

\[
D_{k'|k}(x) = D_{k'|k}(x)B_{k'|k}(x) (10)
\]

where the backward correction term is given by

\[
B_{k'|k}(x) = 1 - p_S(x) + p_S(x) \int \frac{D_{k+1|k}(y)f_{k+1|k}(y|x)}{D_{k+1|k}(y)} \delta_y .
\]

III. SEQUENTIAL MONTE CARLO PHD FILTERING AND SMOOTHING

SMC implementations of the PHD filter and smoother are necessary when dealing with nonlinear models, for example, when the observations are generated according to a range-bearing observation model. This section provides an overview of the SMC formulations of the filter and smoother which are used for subsequent simulations.

A. SMC PHD filter

In the SMC, or particle, PHD filter, the intensity at time \( k-1 \) is approximated using a set of \( N_{k-1} \) weighted samples as

\[
D_{k-1|k-1}(x) = \sum_{i=1}^{N_{k-1}} w_{i|k-1}(x) \delta_{i|k-1}(x) (12)
\]

where \( w_{i|k-1}(x) \) indicates the weight associated with the state \( x_{i|k-1} \) and \( \delta_{i|k-1}(x) \) denotes the Dirac delta function.

The particle PHD is approximated using "persistent" and "new-born" particles [16], where the persistent particles correspond to targets that are currently being tracked. For the particle filter importance density, persistent particles for the existing targets are drawn from the sampling distribution

\[
x_{i,\gamma,k-1|k-1} \sim q(x_{i,\gamma,k-1|k-1} | \gamma_{k-1} \mathcal{Z}_k). (13)
\]

For example, this distribution can be chosen to be the transition density \( f_{k|k-1}(x_{i,\gamma,k-1|k-1}) \). Particles for the birth term are drawn from a birth process as

\[
x_{i,\gamma,k-1|k-1} \sim b(x|\mathcal{Z}_k). (14)
\]

In [16] and [17], it was proposed that the birth process was driven by the measurements with Gaussian components centred around the observations. The predicted PHD is expressed as [16], [18]

\[
D_{k|k-1}(x) = \sum_{i=1}^{N_{k-1}} w_{i|k-1}(x) \delta_{i|k-1}(x) (15)
\]

respectively.

The predicted PHD is then updated according to the equations [16], [18]

\[
D_{k|k}(x) = \sum_{i=1}^{N_{k-1}} w_{i|k-1}(x) \delta_{i|k-1}(x) (16)
\]
B. SMC Forward-Backward PHD Smoothing

The smoothed weights are given by the expression [9]

\[ D_{k' | k}(x) = \sum_{j=1}^{N_{k'}} w_{k' | k}^{(j)} \delta_{x_{k'}^{(j)}}(x) \]  \hspace{1cm} (17)

The smoothed weights are given by the expression [9]

\[ w_{k' | k}^{(i)} = w_{k' | k}^{(i)} (1 - p_{S, k' + 1 | k'}(x_{k'}^{(i)})) \]
\[ + \sum_{j=1}^{N_{k' + 1}} w_{k' + 1 | k'}^{(j)} \gamma_{k' + 1 | k'}(x_{k'}^{(j)}) \] \hspace{1cm} (18)

where

\[ w_{k' + 1 | k'}^{(j)} = w_{k' + 1 | k'}^{(j)} p_{S, k' + 1 | k'}(x_{k'}^{(j)}) f_{k' + 1 | k'}(x_{k'}^{(j)} | x_{k'}^{(i)}) \] \hspace{1cm} (19)

It is noted that new particles are not produced and the recursion merely reweights the particles [9]. This reweighting is expensive due to the evaluation of \( w_{k' + 1 | k'}^{(j)} \) \( \forall j \in \{1, ..., N_{k' + 1}\} \) in equation (18) which is required to correct the weight of a single particle. In high clutter scenarios or when dealing with a large number of targets, the particle smoother becomes highly expensive. The following section considers a variation to the standard particle PHD filter/smoothers to reduce the complexity of the backward recursion.

IV. Target-State Labelling

A tractable particle PHD smoother can be obtained by reducing the computational complexity of the backward recursion in equation (18). One means of achieving this reduction is by limiting the number of terms involved in the nested summation.

The idea behind reducing tractability to the smoother is centred around performing smoothing on a track-wise basis. Each target is represented by a cluster of particles in the PHD and if it were possible to track individual clusters over time, then it is possible to perform track-based smoothing. This is achieved by using an additional, unique label variable to identify clusters.

A. Label-augmented state space

The notion of tagging components in the PHD has been examined before, for example, in [19] and [20]. In those approaches, however, the tag is not incorporated into the state space and receives a more ad hoc treatment. In the framework suggested here, the augmented state is written as

\[ x_{k' | k}^f = \begin{bmatrix} x_{k' | k} \\ T_{k' | k} \end{bmatrix} \] \hspace{1cm} (20)

where \( \tau_{k' | k} \) denotes the tag or label.

The state transition model is described as

\[ x_{k | k - 1} = f_{k | k - 1}(x_{k - 1 | k - 1}) \] \hspace{1cm} (21)

\[ \tau_{k | k - 1} = \tau_{k - 1 | k - 1} \] \hspace{1cm} (22)

where \( f_{k | k - 1}(\cdot) \) denotes the state transition model.

Placing this within the particle PHD framework involves replacing the state \( x \) with the augmented state \( x' \). Additionally, the importance density for the birth process is rewritten as \( x_{\gamma, k | k - 1} \sim b_k(x' | z_k, l) \), where the birth importance function may be time-varying as indicated by the subscript \( k \).

B. Labelled birth intensity

In practice, as suggested in [16] and [17], the birth intensity is chosen as a sum of Gaussian densities centred around each of the \( n_k \) observed states \( Z_k = \{ z_{k,1}, ..., z_{k,n_k} \} \) and each of these densities is assigned a unique label \( \{ T_{k,1}, ..., T_{k,n_k} \} \). Drawing a new-born sample, \( x_{\gamma, k}^f \), from the birth intensity is equivalent to:

\[ l \sim U[a, b] \] \hspace{1cm} (23)

\[ \tau_{\gamma, k} = T_{k, l} \] \hspace{1cm} (24)

\[ x_{\gamma, k} \sim b(x | z_{k, l}) \] \hspace{1cm} (25)

\[ x_{\gamma, k}^f = \begin{bmatrix} x_{\gamma, k} \\ \tau_{\gamma, k} \end{bmatrix} \] \hspace{1cm} (26)

where \( U[a, b] \) denotes a discrete uniform distribution on the interval \([a, b]\). Thus, each cluster of particles, drawn from a single Gaussian component, will be given the same label at the time of birth. Since the model does not allow any change to the labels, the labels can be used to track the evolution of the cluster over time.

The constraint on the label update specified in equation (22) can be used to significantly reduce the complexity of the backward recursion in the smoother. In particular, since the label does not change over time, the transition probability for two states with different labels will be zero:

\[ f_k | k - 1(x_{k | k - 1}^{(i)} | x_{k - 1 | k - 1}^{(j)}) = 0 \hspace{0.5cm} \forall j \neq i \] \hspace{1cm} (27)

By taking this into account, the number of terms required to evaluate the backward recursion is severely constrained and a tractable smoother is achieved.

C. Computational complexity

The fast particle PHD smoother developed here is centred around a reduction in the complexity of the backward recursion presented in equation (18). Reducing the number of terms required in the nested summation reduces the complexity of the smoother.

If \( M \) particles are assigned to each target, the backward recursion in equation (18) has a complexity of \( O(n^2 M^2) \), where the number of targets is \( n \). The labelled state space introduced in Section IV-A allows for a reduction in the complexity by identifying particles which provide a non-zero contribution and removing the need to evaluate all the terms.

Equation (22) specifies that the label associated with a particle may not change. The non-zero terms present in the backward recursion (equation (18)) correspond to those particles at consecutive time instants that possess the same label. If the \( i^{th} \) particle at time \( k' \) possesses the label \( \tau_k^{(i)} \), then calculating the smoothed weight of this particle using equation (18) involves selecting only those particles at time
which possess the same label, \( s_k^{(i)} \), since the transition probability will be zero for all remaining particles at time \( k' + 1 \).

By adopting this approach, the complexity of the smoother is reduced from quadratic to linear in the number of targets \( O(nM^2) \). It is noted that the complexity of the smoother without labelling is also affected by the number of clutter measurements while the label-based smoother is unaffected by clutter. This is illustrated using an example in Section V-A.

V. Results

This section presents results analysing the performance of the particle smoother using the label-augmented state space. The first set of results examine the impact of labelling on the smoothed intensity. The aim is to illustrate the reduction in complexity achieved through the use of the labelled state space.

The second set of results illustrate an example of smoothing using a nonlinear range-bearing observation model in the presence of a moderate number of targets and moderate to high levels of clutter. The estimation error of the smoothed estimates is compared with that of the filtered estimates and the approach adopted here is prescribed as a suitable means to achieve tractable smoothing in practice.

A. Impact of labelling

This section illustrates the impact of labelling on the computational complexity of the smoother for varying number of target and clutter measurements. In order to present the reduction in complexity in the labelled smoother, the average number of transition probability evaluations, \( f_k[k-1](x_{k|k-1}|x_{k-1|k-1}) \), required to evaluate equation (18) is considered while varying the number of targets and clutter in the scene. Plotting the number of evaluations versus target number and clutter intensity shows how the cost of the smoother scales with the average number of observations.

A measurement-driven birth intensity is used to initialise new targets [16], [17] and the probability of detection is set as \( p_D = 0.98 \). The number of particles per new-born target is set to 500, and 1000 particles are equally divided between missed detection and detection components for each target.

Figure 1 illustrates the number of transition probability function evaluations required for a smoother with lag equal to 1. It is seen that the complexity grows quadratically with increasing number of targets when labelling is not used. In contrast, the use of labelling results in a smoother with complexity that is linear in the number of targets.

Figure 2 illustrates the increase in complexity as the clutter intensity \( \lambda_k \) increases. In the absence of labelling, complexity grows linearly while label-based smoothing is seen to have constant complexity (for a given number of targets). Figure 3 shows the speed-up of the labelled smoother, calculated as the ratio of the number of function evaluations required by the full and label-based smoothers.

B. Smoothing in moderate and high clutter environments

It has been stressed that the main aim of labelling is to make the backward recursion tractable in the presence of a high number of targets and clutter. To this end, such an example is presented here.

The scenario involves a maximum of 12 targets with an average of 30 clutter measurements per time step. Figure 4 illustrates the true target locations in Cartesian co-ordinates (solid lines) as well as an example of received target measurements and clutter observations.

A constant velocity model is used for the targets while the observations are modelled using a nonlinear range-bearing model with zero mean additive white Gaussian noise. The single target state is specified as the target in Cartesian co-ordinates with the associated velocity \( \dot{x}_k = [p_x, v_x, p_y, v_y]^T \) and the measurement is given by the range and angle of arrival \( z_k = [r_k, \theta_k]^T \). The state transition
Figure 2. Number of transition probability function evaluations in the backward recursion for varying number of targets and clutter measurements as the number of targets increases from 1 to 6 when (a) labelling is not used and (b) labelling is used. The number of particles used for new-born, missed and detected target clusters is 500 each and the probability of detection is 0.98.

Figure 3. Speed-up achieved in the backward recursion when labelling is used for varying number of clutter measurements as the number of targets increases from 1 (bottom line) to 6 (top line).

Figure 4. $x$ and $y$ co-ordinates of the targets and observations. Solid lines denote true target tracks.

and observation models are given by the equations:

$$x_k = \begin{bmatrix} 1 & \Delta T & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & \Delta T \\ 0 & 0 & 0 & 1 \end{bmatrix} x_{k-1} + \begin{bmatrix} \Delta T^2 & 0 \\ 0 & \Delta T^2 \\ 0 & 0 \end{bmatrix} w_{k-1}$$ (28)

$$z = \begin{bmatrix} \sqrt{p_{x,k}^2 + p_{y,k}^2} \\ \text{arctan} \left( \frac{p_{y,k}}{p_{x,k}} \right) \end{bmatrix} x_k + v_k$$ (29)

where the time between successive observations is given by $\Delta T = 1$, $w_{k-1} \sim \mathcal{N}(0, \sigma_w^2 I)$ is the process noise and the observation noise is $v_k \sim \mathcal{N}(0, \text{diag}([\sigma_r^2, \sigma_\theta^2]^T))$. The variance on the range is $\sigma_r^2 = 3$ m and the variance on the angle of arrival is $\sigma_\theta^2 = 0.035$ rad. The clutter follows a Poisson distribution with intensity $\lambda_c = 2.4 \times 10^{-3}$ (rad m$^{-1}$) over the region $[-\pi, \pi]$rad $\times [0, 2000]$ m which results in an average of 30 clutter measurements at each time instant.

Newborn targets are centred around received observations [16], [17] and each new cluster is initialised using 250 particles. During resampling, clusters arising from misdetections are assigned 250 particles, while updated clusters are assigned 500 particles. The smoother is run with a fixed lag of 3 time steps and the OSPA error [21] is averaged over 100 Monte Carlo runs.

Figures 5 and 6 show the average OSPA cardinality and localisation error for the filter and smoother using parameters cut-off $c = 100$ and power $p = 1$ so that the total OSPA error is the sum of the localisation and cardinality error. It is seen that the cardinality error of the filter and smoother are approximately the same except at times of target death. This effect has been explained in [9] where targets in the smoother suffer a premature death with a lag equal to the smoothing lag. The localisation error is reduced in the case of the smoother as seen in Figure 6. These results are obtained in the presence of moderate number of targets as well as moderate levels of clutter and illustrate the benefits of using label-based smoothing.
VI. CONCLUSIONS

The particle PHD smoother as derived in [9] exhibits computational complexity which is quadratic in the number of targets. This makes the smoother impractical in the presence of a large number of targets and in high clutter.

This paper has proposed augmenting the particle state space with a label to reduce the complexity of the smoother. The label is used to track the evolution of particle clusters so that the backward recursion is only applied to clusters which possess the same label at consecutive time instants. The new label-based smoother exhibits computational complexity that is linear in the number of targets and independent of the clutter intensity which allows for tractable smoothing of a large number of targets.

ACKNOWLEDGEMENTS

Daniel Clark and Sharad Nagappa are members of the Joint Research Institute in Signal and Image Processing within the Edinburgh Research Partnership. This work was supported by the Engineering and Physical Sciences Research Council under grant EP/H010866/1. Daniel Clark is a Royal Academy of Engineering / EPSRC Research Fellow.

REFERENCES

