Closed Form Solutions to Forward-Backward Smoothing

Ba-Ngu Vo, Ba-Tuong Vo, and Ronald P. S. Mahler

***Abstract*— We propose a closed form Gaussian sum smoother and, more importantly, closed form smoothing solutions for increasingly complex problems arising from practice, including tracking in clutter, joint detection and tracking (in clutter), and multiple target tracking (in clutter) via the Probability Hypothesis Density. The solutions are based on the corresponding forward-backward smoothing recursions that involve forward propagation of the filtering densities, followed by backward propagation of the smoothed densities. The key to the exact solutions is the use of alternative forms of the backward propagations, together with standard Gaussian identities. Simulations are also presented to verify the proposed solutions.**

***Index Terms*— Smoothing, Filtering, Gaussian Sum Smoother, PHD, Bernoulli model, target tracking.**

I. Introduction

This paper considers the problem of smoothing for state space models. These models have attracted considerable research interest for several decades, spanning diverse disciplines from statistics, engineering, to econometrics [2], [36]. Smoothing together with filtering and prediction are three important interrelated problems in state space estimation, which essentially amount to calculating

$$p_{k|l}(x_k|z_{1:l}) \tag{1}$$

the probability density of the state x_k at time k given the observation history $z_{1:l} = (z_1, \ldots, z_l)$ up to time l. Smoothing, filtering and prediction, respectively, refer to the cases $l > k$, $l = k$, and $l < k$. In filtering the objective is to recursively estimate the current state given the observation history up to the current time k. Smoothing can yield significantly better estimates than filtering by delaying the decision time (k) and using data at a later time ($l > k$) [24], [12].

Analytic filtering solutions such as the Kalman filter and Gaussian sum filter, for linear Gaussian and linear Gaussian mixture models, have opened up numerous research avenues and pervaded many application areas [14], [27], [1], [2]. For general non-linear models, Sequential Monte Carlo (SMC) or particle filters have recently emerged as powerful numerical approximations [11], [17], [6], [7].

Research in smoothing has experienced similar developments as in filtering, except for the smoothing analogue of the Gaussian sum filter. For general non-linear models, SMC approximations have been proposed for various smoothing schemes including *smoothing-while-filtering* [17], *forward-backward smoothing* [13], [6], [10], (generalized) *two-filter smoothing* [5], [8], and *block-based smoothing* [9]. For the special case of linear Gaussian models, an analytic smoothing solution exists in the form of the Kalman smoother [2]. However, for linear Gaussian mixture models, the Gaussian sum smoother–the smoothing analogue of the Gaussian sum filter–still remains elusive.

Like the Gaussian sum filter, the Gaussian sum smoother is of great practical importance. Even with linear Gaussian transition kernel and likelihood function, the filtering density is inherently non-Gaussian if the initial prior is non-Gaussian. In time series analysis, linear Gaussian models are not adequate for handling outliers or abrupt changes in structure [15], [18]. An analytic smoothing solution for linear Gaussian mixture model opens up new analytical approximations to non-Gaussian state-space smoothing since Gaussian mixtures can, in principle, approximate any density [19].

In many practical problems, such as target tracking, the standard linear Gaussian mixture model is not adequate and more sophisticated models are required. One of the most basic problems is tracking in clutter, where the filtering density is inherently a Gaussian mixture even with a Gaussian initial prior and linear Gaussian transition kernel and likelihood function [3], [21]. Such a problem requires a *(state space) model with finite set observations* [21], [31]. Another problem is joint detection and tracking (in clutter), where the target of interest may not always be present and exact knowledge of target existence cannot be determined from observations due to clutter and detection uncertainty [21], [32]. A *Bernoulli (state space) model* (with finite set observations) is required to accommodate presence and absence of the target [32], [34]. Yet another basic problem, but far more challenging, is multiple target tracking, where the number of targets varies randomly in time, obscured by clutter, detection uncertainty and data association uncertainty. Recently, a *forward-backward Probability Hypothesis Density (PHD) recursion* has been proposed for multiple target smoothing [25], [23]. The PHD is intrinsically multi-modal, indeed, the filtering PHD is a Gaussian mixture even under linear Gaussian multi-target assumptions [29]. Smoothing for these models has a wide range of applications, and closed form solutions offer a versatile set of tools.

Acknowledgement: This work is supported by the Australian Research Council under the Future Fellowship FT0991854 and Discovery Early Career grant DE120102388
B.-N. Vo and B.-T. Vo are with the School of Electrical, Electronic and Computer Engineering, The University of Western Australia, Crawley, WA 6009, Australia (email: ba-ngu.vo@uwa.edu.au, ba-tuong.vo@uwa.edu.au)
R. P. S. Mahler is with the Advanced Technology Group, Lockheed Martin MS2 Tactical Systems, Eagan, Minnesota (email: ronald.p.mahler@lmco.com)

In this paper, we propose a Gaussian sum smoother and, more importantly, analytic smoothing solutions for: state space models with finite set observations; Bernoulli state space models; and the PHD. Our solutions are based on corresponding forward-backward smoothing recursions, each consisting of a forward pass that propagates the filtering density forward to time l, followed by a backward pass that propagates the smoothed density backward to time $k < l$. The forward passes can be accomplished analytically via the corresponding filters, henceforth, our contributions are exact solutions to the backward smoothing recursions for:

- linear Gaussian mixture models,
- linear Gaussian models with finite set observations,
- linear Gaussian Bernoulli models,
- PHD under linear Gaussian multi-target assumptions[1].

The key to these solutions is the application of standard Gaussian identities to novel alternative forms of the backward smoothing recursions.

The proposed Gaussian sum smoother is detailed in Section II while the closed form smoothing solutions for models with finite set observations, Bernoulli models, and the PHD are detailed in Section III. In Section IV, we derive the canonical solutions to the backward smoothing equations for these models. Numerical illustrations are presented in Section V.

II. The Gaussian sum smoother

Brief reviews of forward-backward smoothing and the linear Gaussian mixture model are provided in subsection II-A. In subsection II-B we present an alternative form of the backward recursion, key to the development of our closed form solution. For clarity of presentation, the Gaussian sum smoother is developed for a simpler case first in subsection II-C, while the full Gaussian sum smoother is presented in subsection II-D.

A. Forward-Backward Smoothing

In a state space model, the state of the system follows a Markov process on the state space $\mathcal{X}$, with initial prior p_0 and *transition kernel* $f_{k|k-1}(\cdot|\cdot)$, i.e. given a state x_{k-1} at time $k-1$, the probability density of a transition to the state x_k at time k is

$$f_{k|k-1}(x_k|x_{k-1}). \tag{2}$$

This Markov process is partially observed in the observation space $\mathcal{Z}$ as modeled by the *likelihood function* $g_k(\cdot|\cdot)$, i.e. given a state x_k at time k, the probability density of receiving the observation $z_k \in \mathcal{Z}$ is

$$g_k(z_k|x_k). \tag{3}$$

Forward-backward smoothing consists of a forward pass that propagates the filtering density forward to time l, followed by a backward pass that propagates the smoothing density backward to time $k < l$ (see for example [15]). More concisely, using the standard inner product notation $\langle f, g\rangle = \int f(\zeta)g(\zeta)d\zeta$, the forward-backward smoothing recursion consists of prediction, update and backward smoothing given respectively by

$$p_{k|k-1}(x) = \langle p_{k-1|k-1}, f_{k|k-1}(x|\cdot)\rangle \tag{4}$$

$$p_{k|k}(x) = \frac{g_k(z_k|x)p_{k|k-1}(x)}{\langle g_k(z_k|\cdot), p_{k|k-1}\rangle} \tag{5}$$

$$p_{k-1|l}(x) = p_{k-1|k-1}(x)\left\langle \frac{p_{k|l}}{p_{k|k-1}}, f_{k|k-1}(\cdot|x)\right\rangle \tag{6}$$

Note that for notational compactness we omitted the dependencies on the data in the prediction, filtering and smoothing densities.

In the forward pass, starting with $p_{k|k}$ the prediction densities $p_{k+1|k}, ..., p_{l|l-1}$ and the filtering densities $p_{k+1|k+1}, ...,$ $p_{l|l}$ are computed via the prediction and update recursion. In the backward pass, starting with $p_{l|l}$ the smoothing densities $p_{l-1|l}, ..., p_{k|l}$ are computed via the backward smoothing recursion.

Let $\mathcal{N}(\cdot; m, P)$ denote a Gaussian density with mean m and covariance P, and define

$$\mathcal{N}_{H,R}(z;\zeta) \triangleq \mathcal{N}(z; H\zeta, R)$$

(for appropriate matrices H and R) when we consider $\mathcal{N}(z; H\zeta, R)$ as a function of ζ. The Kalman smoother [2] is a closed form smoothing solution for the *linear Gaussian* (LG) model,

$$p_0(x) = \mathcal{N}(x; m_0, P_0) \tag{7}$$

$$f_{k|k-1}(\zeta|x) = \mathcal{N}_{F_{k|k-1},Q_k}(\zeta; x) \tag{8}$$

$$g_k(z|\zeta) = \mathcal{N}_{H_k,R_k}(z;\zeta) \tag{9}$$

where, m_0, P_0, $F_{k|k-1}$, Q_k, H_k, and R_k are given model parameters.

In a *linear Gaussian mixture* (LGM) model,

$$p_0(x) = \sum_{i=1}^{J_0} w_0^{(i)}\mathcal{N}(x; m_0^{(i)}, P_0^{(i)}) \tag{10}$$

$$f_{k|k-1}(\zeta|x) = \sum_{i=1}^{J_{f,k|k-1}} w_{f,k|k-1}^{(i)}\mathcal{N}_{F_{k|k-1}^{(i)},Q_k^{(i)}}(\zeta; x) \tag{11}$$

$$g_k(z|\zeta) = \sum_{i=1}^{J_{g,k}} w_{g,k}^{(i)}\mathcal{N}_{H_k^{(i)},R_k^{(i)}}(z;\zeta), \tag{12}$$

where $m_0^{(i)}$, $P_0^{(i)}, i = 1, ..., J_0$, $F_{k|k-1}^{(i)}$, $Q_k^{(i)}$, $i = 1, ..., J_{f,k|k-1}$ and $H_k^{(i)}$, $R_k^{(i)}$, $i = 1, ..., J_{g,k}$ are given model parameters. Under LGM assumption, a closed form filtering solution is the Gaussian sum filter [27], [1], which recursively propagates forward the (Gaussian mixture) predicted and filtered densities:

$$p_{k|k-1}(x) = \sum_{i=1}^{J_{k|k-1}} w_{k|k-1}^{(i)}\mathcal{N}(x; m_{k|k-1}^{(i)}, P_{k|k-1}^{(i)}), \tag{13}$$

$$p_{k|k}(x) = \sum_{i=1}^{J_{k|k}} w_{k|k}^{(i)}\mathcal{N}(x; m_{k|k}^{(i)}, P_{k|k}^{(i)}), \tag{14}$$

However, for smoothing, a closed form solution has not yet been found, since the quotient of two Gaussian mixtures in (6) is not necessarily a Gaussian mixture (even if the transition

[1]Preliminary results for the PHD smoother have been announced in the conference paper [33].

density and likelihood are linear Gaussian), see for example [16] pp. 609.

B. Backward corrector recursion

To facilitate the derivation of the closed form smoothing solution, we rewrite the update and backward smoothing recursions (5) and (6) in the following form:

$$p_{k|k}(x) = L_k(z_k; x) p_{k|k-1}(x) \quad (15)$$

$$p_{k|l}(x) = p_{k|k}(x) B_{k|l}(x) \quad (16)$$

where

$$L_k(z_k; x) = \frac{g_k(z_k|x)}{\langle g_k(z_k|\cdot), p_{k|k-1}\rangle} \quad (17)$$

$$B_{k|l}(x) = \left\langle \frac{p_{k+1|l}}{p_{k+1|k}}, f_{k+1|k}(\cdot|x)\right\rangle \quad (18)$$

with $B_{l|l}(\cdot) = 1$. The key to solving the backward smoothing problem is to note that the so-called backward corrector $B_{k|l}$ can be recursively computed as follows.

Proposition 1: For $k \leq l$,

$$B_{k-1|l}(x) = \langle B_{k|l} L_k(z_k; \cdot), f_{k|k-1}(\cdot|x)\rangle \quad (19)$$

Proof: It follows from (16) and (15) that

$$\frac{p_{k|l}}{p_{k|k-1}} = \frac{p_{k|k} B_{k|l}}{p_{k|k-1}} = L_k(z_k; \cdot) B_{k|l},$$

which upon substitution into (18) with k replaced by $k-1$, gives (19). □

The backward corrector recursion (19) resembles the information filter in the two-filter smoother of [4], [16]. However, it does not have an information filter interpretation since the backward corrector is not a probability density. In fact, a filter interpretation is not necessary for the derivation of a closed form solution to the backward corrector recursion. Next, we use this backward corrector recursion to derive closed form expressions for the backward corrector $B_{k|l}$ and subsequently the smoothed density $p_{k|l}$.

C. Smoothing for linear Gaussian model with Gaussian mixture prior

The closed form solution to the backward corrector recursion (19) is most easily seen via a simpler special case of the LGM model. Specifically, the dynamic and measurement models are assumed linear Gaussian, but the prior is a Gaussian mixture and consequently the prediction and filtered densities are Gaussian mixtures. The following proposition and its corollary provide closed form expressions for the backward corrector $B_{k|l}$ and the smoothed density $p_{k|l}$ respectively.

Proposition 2: Under the linear Gaussian dynamic and measurement model (8), (9), suppose that at time k, the backward corrector has the form

$$B_{k|l}(x_k) = \frac{\mathcal{N}_{C_k, D_k}(\zeta_k; x_k)}{r_k} \quad (20)$$

then at time $k-1$ the backward corrector is given by

$$B_{k-1|l}(x_{k-1}) = \frac{\mathcal{N}_{\ddot{C}_k, \ddot{D}_k}([\zeta_k^T, z_k^T]^T; x_{k-1})}{\nu_k(z_k) r_k} \quad (21)$$

where

$$\ddot{C}_k = \begin{bmatrix} C_k \\ H_k \end{bmatrix} F_{k|k-1}, \quad (22)$$

$$\ddot{D}_k = \begin{bmatrix} D_k & 0 \\ 0 & R_k \end{bmatrix} + \begin{bmatrix} C_k \\ H_k \end{bmatrix} Q_k \begin{bmatrix} C_k^T & H_k^T \end{bmatrix} \quad (23)$$

$$\nu_k(z) = \sum_{j=1}^{J_{k|k-1}} w_{k|k-1}^{(j)} \mathcal{N}_{H_k, R_k + H_k P_{k|k-1}^{(j)} H_k^T}(z; m_{k|k-1}^{(j)}) \quad (24)$$

Proof: Since the measurement model is linear Gaussian (9), and prediction density is a Gaussian mixture of the form (13), we have

$$\begin{aligned} &\langle g_k(z|\cdot), p_{k|k-1}\rangle \\ &= \sum_{j=1}^{J_{k|k-1}} w_{k|k-1}^{(j)} \left\langle \mathcal{N}_{H_k, R_k}(z; \cdot), \mathcal{N}(\cdot; m_{k|k-1}^{(j)}, P_{k|k-1}^{(j)})\right\rangle \\ &= \nu_k(z) \end{aligned} \quad (25)$$

by virtue of the convolution of Gaussians in Lemma 18 (Appendix VII-A). Hence,

$$\begin{aligned} B_{k|l}(x_k) L_k(z_k; x_k) &= \frac{\mathcal{N}_{C_k, D_k}(\zeta_k; x_k)}{r_k} \frac{\mathcal{N}_{H_k, R_k}(z_k; x_k)}{\nu_k(z_k)} \\ &= \frac{\mathcal{N}_{\tilde{C}_k, \tilde{D}_k}([\zeta_k^T, z_k^T]^T; x_k)}{r_k \nu_k(z_k)} \end{aligned}$$

where

$$\tilde{C}_k = \begin{bmatrix} C_k \\ H_k \end{bmatrix}, \quad \tilde{D}_k = \begin{bmatrix} D_k & 0 \\ 0 & R_k \end{bmatrix}.$$

Using recursion (19) from Proposition 1,

$$\begin{aligned} B_{k-1|l}(x_{k-1}) &= \langle B_{k|l} L_k(z_k; \cdot), f_{k|k-1}(\cdot|x_{k-1})\rangle \\ &= \left\langle \frac{\mathcal{N}_{\tilde{C}_k, \tilde{D}_k}([\zeta_k^T, z_k^T]^T; \cdot)}{r_k \nu_k(z_k)}, \mathcal{N}_{F_{k|k-1}, Q_k}(\cdot; x_{k-1})\right\rangle, \end{aligned}$$

and then using the convolution of Gaussians in Lemma 18 again gives (21), (22). □

Remark: The premise of Proposition 2 is that the backward corrector at time k has the form (20), i.e. Gaussian in some linear transformation of x_k. Using the convention $\mathcal{N}_{[],[]}(\zeta; x) \triangleq 1$, where [] is the MATLAB notation for the null matrix satisfying

$$\begin{bmatrix} [] \\ H \end{bmatrix} = H,$$

and noting that the backward corrector iteration starts with $B_{l|l} = 1$, it is clear that the premise of the above Proposition holds for $k = l$. Consequently, it follows by induction from Proposition 2 that all subsequent backward correctors are Gaussians in some linear transformations of the state vector. This result allows the smoothed density to be written as a Gaussian mixture by using a standard result on products of Gaussians.

Corollary 3: Under the linear Gaussian dynamic and measurement model (8), (9), the smoothed density $p_{k|l}$ is a Gaussian mixture

$$p_{k|l}(x) = \frac{1}{r_k} \sum_{i=1}^{J_{k|k}} w_{k|k}^{(i)} q_{k|k}^{(i)}(\zeta_k) \mathcal{N}(x; \tilde{m}_{k|k}^{(i)}(\zeta_k), \tilde{P}_{k|k}^{(i)}) \quad (26)$$

where

$$q^{(i)}_{k|k}(\zeta_k) = \mathcal{N}_{C_k, D_k + C_k P^{(i)}_{k|k} C_k^T}(\zeta_k; m^{(i)}_{k|k}) \quad (27)$$

$$\tilde{m}^{(i)}_{k|k}(\zeta_k) = m^{(i)}_{k|k} + K^{(i)}_{k|k}(\zeta_k - C_k m^{(i)}_{k|k}) \quad (28)$$

$$\tilde{P}^{(i)}_{k|k} = (I - K^{(i)}_{k|k} C_k) P^{(i)}_{k|k} \quad (29)$$

$$K^{(i)}_{k|k} = P^{(i)}_{k|k} C_k^T (C_k P^{(i)}_{k|k} C_k^T + D_k)^{-1} \quad (30)$$

and r_k, ζ_k, C_k, D_k are the parameters of the backward corrector $B_{k|l}$.

Proof: Using (16), (14), Proposition 2, and then the Gaussian identity in Lemma 19 (Appendix VII-A), the smoothed density is

$$\begin{aligned} p_{k|l}(x) &= p_{k|k}(x) B_{k|l}(x) \\ &= \sum_{i=1}^{J_{k|k}} \frac{w^{(i)}_{k|k}}{r_k} \mathcal{N}(x; m^{(i)}_{k|k}, P^{(i)}_{k|k}) \mathcal{N}_{C_k, D_k}(\zeta_k; x), \\ &= \frac{1}{r_k} \sum_{i=1}^{J_{k|k}} w^{(i)}_{k|k} q^{(i)}_{k|k}(\zeta_k) \mathcal{N}(x; \tilde{m}^{(i)}_{k|k}(\zeta_k), \tilde{P}^{(i)}_{k|k}). \square \end{aligned}$$

Remark: The normalising constant r_k (which is a function of the measurements $z_{l:k+1}$) is included for completeness, but in practice it is not necessary. Instead we normalise the weights $\{w^{(i)}_{k|k} q^{(i)}_{k|k}(\zeta_k)\}_{i=1}^{J_{k|k}}$. Further, it is not necessary to compute the filtering densities for time $k+1$ to l. Instead, we only need (the Gaussians components of) the filtered density $p_{k|k}$, and predicted density $p_{k+1|k}$ because the backward corrector can be computed without requiring the filtering densities for time $k+1$ to l (see also section IV-B which gives a full non recursive expression for $B_{k|l}$).

The solution presented bears some resemblance to the approximation proposed in [16]. The key difference is that [16] uses a two-filter smoother in which the backward recursion is approximated by a Kalman filter. Specifically, the approach in [16] requires $\tilde{C}_k$ is to be invertible, which is not valid in general (it is not even a square matrix). Our approach does not treat the backward propagation as an information filter nor assume $\tilde{C}_k$, to be invertible.

D. Smoothing for Linear Gaussian Mixture models

We now present the full Gaussian sum smoother by extending Proposition 2 to the LGM model. In this case, the smoothed density is a Gaussian mixture and closed form solutions for the backward corrector $B_{k|l}$ and the smoothed density $p_{k|l}$ are given respectively by the following proposition and its corollary. The proofs are omitted because they are almost identical to that of Proposition 2 and Corollary 3.

Proposition 4: Under the linear Gaussian mixture dynamic and measurement model (11), (12), suppose that at time k, the backward corrector has the form

$$B_{k|l}(x_k) = \sum_{h=1}^{J_{k|l}} w_k^{(h)} \mathcal{N}_{C_k^{(h)}, D_k^{(h)}}(\zeta_k; x_k) \quad (31)$$

then at time $k-1$ the backward corrector is given by

$$B_{k-1|l}(x_{k-1}) = \sum_{h=1}^{J_{k|l}} \sum_{i=1}^{J_{g,k}} \sum_{j=1}^{J_{f,k|k-1}} \frac{w^{(i)}_{g,k} w^{(j)}_{f,k|k-1} w^{(h)}_k}{\sum_{n=1}^{J_{g,k}} w^{(n)}_{g,k} \nu^{(n)}_k(z_k)} \times \mathcal{N}_{\ddot{C}_k^{(h,i,j)}, \ddot{D}_k^{(h,i,j)}}([\zeta_k^T, z_k^T]^T; x_{k-1}) \quad (32)$$

where

$$\ddot{C}_k^{(h,i,j)} = \begin{bmatrix} C_k^{(h)} \\ H_k^{(i)} \end{bmatrix} F^{(j)}_{k|k-1}, \quad (33)$$

$$\ddot{D}_k^{(h,i,j)} = \begin{bmatrix} D_k^{(h)} & 0 \\ 0 & R_k^{(i)} \end{bmatrix} + \begin{bmatrix} C_k^{(h)} \\ H_k^{(i)} \end{bmatrix} Q_k^{(j)} \begin{bmatrix} C_k^{(h)T} & H_k^{(i)T} \end{bmatrix}, \quad (34)$$

$$\nu_k^{(i)}(z) = \sum_{j=1}^{J_{k|k-1}} w^{(j)}_{k|k-1} \mathcal{N}_{H_k^{(i)}, R_k^{(i)} + H_k^{(i)} P^{(j)}_{k|k-1} H_k^{(i)T}}(z; m^{(j)}_{k|k-1}) \quad (35)$$

Remark: The premise of Proposition 4 is that the backward corrector at time k has the form (31), i.e. a Gaussian mixture in some linear transformation of x_k. Noting that the backward corrector iteration starts with $B_{l|l} = \mathcal{N}_{[],[]}$, it is clear that the premise of the above Proposition holds for $k = l$. Consequently, it follows by induction from Proposition 4 that all subsequent backward correctors are Gaussian mixtures in some linear transformations of the state vector. This result allows the smoothed density to be written as a Gaussian mixture.

Corollary 5: Under the linear Gaussian mixture dynamic and measurement model (11), (12), the smoothed density $p_{k|l}$ is a Gaussian mixture

$$p_{k|l}(x) = \sum_{h=1}^{J_{k|l}} \sum_{i=1}^{J_{k|k}} w^{(i)}_{k|k} q^{(i,h)}_{k|k}(\zeta_k) w^{(h)}_k \mathcal{N}(x; \tilde{m}^{(i,h)}_{k|k}(\zeta_k), \tilde{P}^{(i,h)}_{k|k}) \quad (36)$$

where

$$q^{(i,h)}_{k|k}(\zeta_k) = \mathcal{N}_{C_k^{(h)}, D_k^{(h)} + C_k^{(h)} P^{(i)}_{k|k} C_k^{(h)T}}(\zeta_k; m^{(i)}_{k|k}) \quad (37)$$

$$\tilde{m}^{(i,h)}_{k|k}(\zeta_k) = m^{(i)}_{k|k} + K^{(i,h)}_{k|k}(\zeta_k - C_k^{(h)} m^{(i)}_{k|k}) \quad (38)$$

$$\tilde{P}^{(i,h)}_{k|k} = (I - K^{(i,h)}_{k|k} C_k^{(h)}) P^{(i)}_{k|k} \quad (39)$$

$$K^{(i,h)}_{k|k} = P^{(i)}_{k|k} C_k^{(h)T} (C_k^{(h)} P^{(i)}_{k|k} C_k^{(h)T} + D_k^{(h)})^{-1} \quad (40)$$

and ζ_k, $(w_k^{(h)}, C_k^{(h)}, D_k^{(h)})_{h=1}^{J_{k|l}}$ are the parameters of the backward corrector $B_{k|l}$.

Similar to the previous result, in practice there is no need to calculate $\nu^{(i)}_{k+1}(z_{k+1})$ at all. Instead we normalise the weights after multiplying the backward corrector with the filtered density. It is also not necessary to compute the filtering densities for times $k+1$ to l. Instead, we only need (the Gaussians components of) the filtered density $p_{k|k}$, and predicted density $p_{k+1|k}$.

III. Gaussian Mixture Smoothing for Generalized Models

While the Gaussian sum smoother considered in the previous section spans a wide range of applications, it is not adequate for a larger class of practical problems which require more sophisticated models such as state space model with finite set observations, Bernoulli model, and PHD model. In subsection III-A we derive a closed form solution to a generic backward recursion that covers these models. The forward-backward smoothing recursions and their respective closed solutions are then given in subsections III-B, III-C for the finite set observations model, subsections III-D, III-E for the Bernoulli model, and subsections III-F, III-G for the PHD.

A. Generic Gaussian mixture backward propagation

This subsection presents a closed form solution to a generic backward recursion that covers smoothing for models with finite set observation, Bernoulli models, and the PHD. It is shown in subsequent subsections that these backward recursions can all be recasted in terms of the backward corrector $B_{k|l}$ that obeys the following generic backward corrector recursion (starting with $B_{l|l} = 1$)

$$B_{k-1|l}(x) = q_k + p_k \left\langle B_{k|l} L_k(Z_k; \cdot), \mathcal{N}_{F_{k|k-1}, Q_k}(\cdot; x) \right\rangle \quad (41)$$

$$L_k(Z_k; x) = \alpha_k + \sum_{z \in Z_k} w_k(z) \mathcal{N}_{H_k, R_k}(z; x) \quad (42)$$

where q_k, p_k, α_k and $w_k(z)$ are real numbers.

Remark: For clarity of presentation, we assume a linear Gaussian transition kernel. Nonetheless the result can be easily extended to linear Gaussian mixture transition kernel, albeit through a rather cumbersome process.

Proposition 6: Suppose that at time k, the backward corrector has the form

$$B_{k|l}(x) = \sum_{i=1}^{J_{k|l}} w_k^{(i)} \mathcal{N}_{C_k^{(i)}, D_k^{(i)}}(\zeta_k^{(i)}; x), \quad (43)$$

then applying the generic backward corrector recursion (41), (42), yields

$$B_{k-1|l}(x) = q_k + p_k \left(\alpha_k \dot{B}_{k|l}(x) + \sum_{z \in Z} w_k(z) \ddot{B}_{k|l}(x; z) \right) \quad (44)$$

where

$$\dot{B}_{k|l}(x) = \sum_{i=1}^{J_{k|l}} w_k^{(i)} \mathcal{N}_{\dot{C}_k^{(i)}, \dot{D}_k^{(i)}}(\zeta_k^{(i)}; x) \quad (45)$$

$$\dot{C}_k^{(i)} = C_k^{(i)} F_{k|k-1}, \quad (46)$$

$$\dot{D}_k^{(i)} = D_k^{(i)} + C_k^{(i)} Q_k C_k^{(i)T}, \quad (47)$$

$$\ddot{B}_{k|l}(x; z) = \sum_{i=1}^{J_{k|l}} w_k^{(i)} \mathcal{N}_{\ddot{C}_k^{(i)}, \ddot{D}_k^{(i)}}([\zeta_k^{(i)T}, z^T]^T; x) \quad (48)$$

$$\ddot{C}_k^{(i)} = \begin{bmatrix} C_k^{(i)} \\ H_k \end{bmatrix} F_{k|k-1}, \quad (49)$$

$$\ddot{D}_k^{(i)} = \begin{bmatrix} D_k^{(i)} & 0 \\ 0 & R_k \end{bmatrix} + \begin{bmatrix} C_k^{(i)} \\ H_k \end{bmatrix} Q_k \begin{bmatrix} C_k^{(i)T} & H_k^T \end{bmatrix} \quad (50)$$

The proof is similar to Proposition 2, and details are given in Appendix VII-B.

Remark: The premise of Proposition 6 is that the backward corrector at time k has the form (43), i.e. a Gaussian mixture in some linear transformation of x_k. This premise holds for $k = l$ since the backward corrector iteration starts with $B_{l|l} = \mathcal{N}_{[],[]}$. Consequently, it follows by induction from Proposition 6 that all subsequent backward correctors are Gaussian mixtures in some linear transformations of the state.

Remark: Unlike Proposition 4, in which all the components in the mixture share a common mean ζ_k, in the above result each component of the mixture has a different mean $\zeta_k^{(i)}$.

B. Linear Gaussian model with finite set observation

In applications such as target tracking, the state-generated observation is further corrupted by clutter and detection uncertainty. As a result, the observation is no longer a vector, but a finite set of vectors with uncertain origin, and a (state space) model with *finite set observations* is needed [21], [31]. In such a model the following additional parameters are used to characterize detection uncertainty and clutter:

$$p_{D,k}(\zeta) = \text{probability of detection of state } \zeta \text{ at time } k \quad (51)$$

$$\kappa_k(z) = \text{intensity of Poisson clutter at time } k \quad (52)$$

The forward-backward smoothing recursion is the same as that of the standard model, i.e. (4)-(6), with the measurements z_k replaced by a finite set Z_k and the following measurement likelihood [21], [31]:

$$g_k(Z_k|\zeta) = \frac{p_{D,k}(\zeta) \sum_{z \in Z_k} \kappa_k^{Z_k - \{z\}} g_k(z|\zeta) + q_{D,k}(\zeta) \kappa_k^{Z_k}}{e^{\langle \kappa_k, 1 \rangle}} \quad (53)$$

where $q_{D,k}(\zeta) = 1 - p_{D,k}(\zeta)$, $h^Z = \prod_{z \in Z} h(z)$. Note that by convention $h^{\emptyset} = 1$, (even if $h = 0$).

The *linear Gaussian finite set observation* (LG-FSO) state space model assumes a Gaussian mixture initial prior, i.e. (10), linear Gaussian transition kernel and measurement likelihood, i.e. (8), (9), and constant probability of detection. Note that the LG model is a special case of the LG-FSO model with $p_{D,k} = 1$ and $\kappa_k = 0$. For the LG-FSO model, the prediction and filtering densities are Gaussian mixtures of the form (13), (14). The closed form filtering solution which recursively propagates forward the Gaussian mixture prediction and filtering densities can be found in [21], [31]. Due to the non-Gaussian nature of the measurement likelihood, the filtering density is inherently a Gaussian mixture. A closed form smoothing solution for LG-FSO models has not been found.

Replacing the transition density (8) and likelihood (9) in the LG-FSO model with (11) and (12), we obtain a linear Gaussian mixture model with FSO.

C. Smoothing for LG-FSO model

The LG-FSO model described in subsection III-B differs from the conventional state space model only in the likelihood function (53). Hence, following the arguments of Proposition 1, the backward smoothing recursion for the LG-FSO model can be rewritten in the form of the generic backward corrector recursion (41)-(42) with $q_k = 0$, $p_k = 1$ and pseudo-likelihood

$$L_k(Z_k; x) = \frac{g_k(Z_k|x)}{\langle g_k(Z_k|\cdot), p_{k|k-1} \rangle} \quad (54)$$

The following result is, thus, a direct consequence of Proposition 6 (for completeness the proof is given in Appendix VII-B).

Corollary 7: Under the LG-FSO model, suppose that at time k the backward corrector has the form (43), then

$$B_{k-1|l}(x) = \frac{q_{D,k}\kappa_k^{Z_k}\dot{B}_{k|l}(x) + p_{D,k}\sum_{z\in Z_k}\kappa_k^{Z_k-\{z\}}\ddot{B}_{k|l}(x;z)}{q_{D,k}\kappa_k^{Z_k} + p_{D,k}\sum_{z\in Z_k}\kappa_k^{Z_k-\{z\}}\nu_k(z)} \quad (55)$$

where $\nu_k(z)$, $\dot{B}_{k|l}(x)$, and $\ddot{B}_{k|l}(x)$ are given respectively by (24) (45), and (48). Moreover, the smoothed density is a Gaussian mixture given by

$$p_{k|l}(x) = \sum_{i=1}^{J_{k|l}}\sum_{j=1}^{J_{k|k}} w_{k|k}^{(j)} w_k^{(i)} q_{k|k}^{(i,j)}(\zeta_k^{(i)})\mathcal{N}(x;\tilde{m}_{k|k}^{(i,j)}(\zeta_k^{(i)}),\tilde{P}_{k|k}^{(i,j)}) \quad (56)$$

where

$$q_{k|k}^{(i,,j)}(\zeta_k^{(i)}) = \mathcal{N}_{C_k^{(i)},D_k^{(i)}+C_k^{(i)}P_{k|k}^{(j)}C_k^{(i)T}}(\zeta_k^{(i)};m_{k|k}^{(j)}) \quad (57)$$

$$\tilde{m}_{k|k}^{(i,j)}(\zeta_k^{(i)}) = m_{k|k}^{(j)} + K_{k|k}^{(i,j)}(\zeta_k^{(i)} - C_k^{(i)}m_{k|k}^{(j)}) \quad (58)$$

$$\tilde{P}_{k|k}^{(i,j)} = (I - K_{k|k}^{(i,j)}C_k^{(i)})P_{k|k}^{(j)} \quad (59)$$

$$K_{k|k}^{(i,j)} = P_{k|k}^{(j)}C_k^{(i)T}(C_k^{(i)}P_{k|k}^{(j)}C_k^{(i)T} + D_k^{(i)})^{-1} \quad (60)$$

Remark: As in the single-measurement case, the normalising constant is included for completeness, in practice there is no need to calculate it at all. Instead we normalise the weights $\{w_{k|k}^{(i)}q_{k|k}^{(i)}(z_{l:k+1})\}_{i=1}^{J_{k|k}}$. It is also not necessary to compute the filtering densities for time $k+1$ to l. Instead, we only need (the Gaussians components of) the filtered density $p_{k|k}$, and predicted density $p_{k+1|k}$.

D. Linear Gaussian Bernoulli model

The previous models assume that the target is always present in the scene. In practice, the target of interest may not always be present and exact knowledge of target existence cannot be determined from observations due to clutter and detection uncertainty [21] [32]. A *Bernoulli (state space) model* (with finite set observations) is a generalisation of the standard model, which accommodates presence and absence of the target, based on random finite set theory [21] [32]. In a Bernoulli model, the probability law can be specified by a pair of parameters (r, p), where r is the existence probability of the target and p is the probability density that describes the state of the target conditioned on its existence.

In addition to the standard state transition density $f_{k|k-1}(\cdot|\cdot)$, the dynamical description of the Bernoulli model includes the following parameters:

$$p_{R,k|k-1} = \text{probability of entry/reentry at time } k \quad (61)$$

$$f_{R,k|k-1}(\zeta) = \text{density of entry/reentry of state } \zeta \text{ at time } k \quad (62)$$

$$p_{S,k|k-1}(x) = \text{probability of survival to time } k \text{ given state } x \text{ at time } k-1 \quad (63)$$

If the target is not in the scene at time $k-1$, it can enter (or re-enter) the scene with probability $p_{R,k|k-1}$ and occupy state ζ with probability density $f_{R,k|k-1}(\zeta)$, or remain absent from the scene with probability $q_{R,k|k-1} = 1 - p_{R,k|k-1}$. On the other hand, if the target exists and has state x, at time $k-1$, it can survive to the next time step with probability $p_{S,k|k-1}(x)$ and evolve to state ζ with probability density $f_{k|k-1}(\zeta|x)$, or disappear with probability $q_{S,k|k-1}(x) = 1 - p_{S,k|k-1}(x)$.

For the measurement model, if the target exists and has state x, then the likelihood of receiving the measurement Z_k is the same as (53). Otherwise all measurements must originate from clutter and the likelihood is then $e^{-\langle\kappa_k,1\rangle}\kappa_k^{Z_k}$.

The forward-backward propagation for the Bernoulli model is more complex than those of the previous models since the existence probability and the density of the state need to be jointly propagated. For each integer $k > 0$, let $r_{k|l}$ denote the existence probability of a target at time k given the observation history $Z_{1:l} = (Z_1, ..., Z_l)$ up to time l. Then, the forward-backward Bernoulli smoothing recursion consists of the following steps:

- Prediction [34]

$$r_{k|k-1} = p_{R,k|k-1}(1 - r_{k-1|k-1}) + r_{k-1|k-1}\left\langle p_{S,k|k-1}, p_{k-1|k-1}\right\rangle \quad (64)$$

$$p_{k|k-1}(\zeta) = \frac{p_{R,k|k-1}(1 - r_{k-1|k-1})f_{R,k|k-1}(\zeta)}{r_{k|k-1}} + \frac{r_{k-1|k-1}\left\langle p_{S,k|k-1}f_{k|k-1}(\zeta|\cdot), p_{k-1|k-1}\right\rangle}{r_{k|k-1}} \quad (65)$$

- Update [34]

$$r_{k|k} = \frac{r_{k|k-1}\left\langle g_k(Z_k|\cdot), p_{k|k-1}\right\rangle}{\frac{(1-r_{k|k-1})\kappa_k^{Z_k}}{e^{\langle\kappa_k,1\rangle}} + r_{k|k-1}\left\langle g_k(Z_k|\cdot), p_{k|k-1}\right\rangle}, \quad (66)$$

$$p_{k|k}(x) = \frac{g_k(Z_k|x)p_{k|k-1}(x)}{\left\langle g_k(Z_k|\cdot), p_{k|k-1}\right\rangle}. \quad (67)$$

- Backward smoothing [35],

$$r_{k-1|l} = 1 - (1 - r_{k-1|k-1}) \times \left[\alpha_{R,k|l} + \beta_{R,k|l}\left\langle\frac{p_{k|l}}{p_{k|k-1}}, f_{R,k|k-1}\right\rangle\right] \quad (68)$$

$$p_{k-1|l}(x) = p_{k-1|k-1}(x)\frac{B_{k-1|l}(x)}{\left\langle B_{k-1|l}(x), p_{k-1|k-1}\right\rangle} \quad (69)$$

$$B_{k-1|l}(x) = \alpha_{S,k|l}(x) + \beta_{S,k|l}(x)\left\langle\frac{p_{k|l}}{p_{k|k-1}}, f_{k|k-1}(\cdot|x)\right\rangle, \quad (70)$$

where

$$\alpha_{R,k|l} = q_{R,k|k-1}\frac{(1 - r_{k|l})}{(1 - r_{k|k-1})},$$

$$\beta_{R,k|l} = p_{R,k|k-1}\frac{r_{k|l}}{r_{k|k-1}},$$

$$\alpha_{S,k|l}(x) = q_{S,k|k-1}(x)\frac{(1 - r_{k|l})}{(1 - r_{k|k-1})},$$

$$\beta_{S,k|l}(x) = p_{S,k|k-1}(x)\frac{r_{k|l}}{r_{k|k-1}}.$$

with $B_{l|l}(\cdot) = 1$.

The *linear Gaussian Bernoulli* (LG-Bernoulli) model assumes: constant detection probability, Gaussian mixture initial prior, linear Gaussian transition kernel and measurement likelihood, as in the LG-FSO model. Further, the LG-Bernoulli model assumes constant probability of survival $p_{S,k|k-1}$, constant probability of entry/reentry $p_{R,k|k-1}$, and Gaussian mixture entry/reentry density, i.e.

$$f_{R,k|k-1}(\zeta) = \sum_{j=1}^{J_{R,k|k-1}} w_{R,k|k-1}^{(j)} \mathcal{N}(\zeta; m_{R,k|k-1}^{(j)}, P_{R,k|k-1}^{(j)}) \quad (71)$$

where $J_{R,k|k-1}$, $w_{R,k|k-1}^{(j)}$, $m_{R,k|k-1}^{(j)}$, $P_{R,k|k-1}^{(j)}$, $j = 1, \ldots, J_{R,k|k-1}$, are given model parameters. Note that the LG-FSO model is a special case of the LG-Bernoulli model with $r_0 = 1$, $p_{S,k|k-1} = 1$, $p_{R,k|k-1} = 0$. For the LG-Bernoulli model, the prediction and filtering densities are Gaussian mixtures of the form (13), (14). The closed form filtering solution which recursively propagates forward the Gaussian mixture prediction and filtering densities can be found in [34].

Replacing the transition density (8) and likelihood (9) in the LG-Bernoulli model with (11) and (12) we have the Gaussian mixture Bernoulli model.

E. The GM-Bernoulli smoother

This subsection presents the Gaussian mixture smoothing solution for the LG-Bernoulli model described in subsection III-D. This model is different from the conventional state space model due to the uncertainty in the existence of the state. The Bernoulli smoother propagates the probability of existence in addition to the probability density.

Similar to the previous smoothing solutions, it is more convenient to rewrite the Bernoulli backward recursion (69) in backward corrector form:

$$p_{k|k}(x) = L_k(Z_k; x) p_{k|k-1}(x) \quad (72)$$

$$p_{k|l}(x) = p_{k|k}(x) \frac{B_{k|l}(x)}{\langle B_{k|l}, p_{k|k} \rangle}, \quad (73)$$

where the pseudo-likelihood $L_k(Z_k; x)$ is given by (54). The following proposition shows that the recursion for the backward corrector falls under the generic backward recursion (41)-(42) with $q_k = \alpha_{S,k|l}$ and $p_k = \beta_{S,k|l} / \langle p_{k|k} B_{k|l} \rangle$.

Proposition 8: For $k \leq l$,

$$B_{k-1|l}(x) = \alpha_{S,k|l} + \frac{\beta_{S,k|l}}{\langle p_{k|k} B_{k|l} \rangle} \langle B_{k|l} L_k(Z_k; \cdot), f_{k|k-1}(\cdot|x) \rangle \quad (74)$$

$$r_{k-1|l} = 1 - (1 - r_{k-1|k-1}) \times \left(\alpha_{R,k|l} + \frac{\beta_{R,k|l}}{\langle p_{k|k} B_{k|l} \rangle} \langle B_{k|l} L_k(Z_k; \cdot), f_{R,k|k-1} \rangle \right) \quad (75)$$

Proof: It follows from (72) and (73) that

$$\frac{p_{k|l}}{p_{k|k-1}} = \frac{p_{k|k} B_{k|l}}{\langle p_{k|k}, B_{k|l} \rangle p_{k|k-1}} = \frac{B_{k|l}}{\langle p_{k|k}, B_{k|l} \rangle} L_k(Z_k; \cdot),$$

which upon substitution into (68) and (70) with k replaced by $k-1$, gives (74) and (75), respectively. □

The closed form solution for $B_{k|l}$ under the Bernoulli model then follows from the generic solution in Proposition 6. The complete closed form backward corrector for the Bernoulli model is given by the following result. The proof is given in Appendix VII-B.

Corollary 9: Under the LG-Bernoulli model, suppose that at time k, the backward corrector has the form (43) then at time $k-1$ the backward corrector and smoothed existence probability are given by

$$B_{k-1|l}(x) = \alpha_{S,k|l} + \frac{\beta_{S,k|l}}{\nu_{k|l}} \times \frac{q_{D,k} \kappa_k^{Z_k} \dot{B}_{k|l}(x) + p_{D,k} \sum_{z \in Z_k} \kappa_k^{Z_k - \{z\}} \ddot{B}_{k|l}(x; z)}{q_{D,k} \kappa_k^{Z_k} + p_{D,k} \sum_{z \in Z_k} \kappa_k^{Z_k - \{z\}} \nu_k(z)},$$

$$\frac{1 - r_{k-1|l}}{1 - r_{k-1|k-1}} = \alpha_{R,k|l} + \frac{\beta_{R,k|l}}{\nu_{k|l}} \times \frac{q_{D,k} \kappa_k^{Z_k} \acute{B}_{k|l} + p_{D,k} \sum_{z \in Z_k} \kappa_k^{Z_k - \{z\}} \H{B}_{k|l}(z)}{q_{D,k} \kappa_k^{Z_k} + p_{D,k} \sum_{z \in Z_k} \kappa_k^{Z_k - \{z\}} \nu_k(z)}$$

where $\nu_k(z)$, $\dot{B}_{k|l}(x)$, $\ddot{B}_{k|l}(x)$ are given respectively by (24), (45), (48) and

$$\nu_{k|l} = \sum_{j=1}^{J_{k|k}} \sum_{i=1}^{J_{k|l}} w_{k|k}^{(j)} w_k^{(i)} \mathcal{N}_{C_k^{(i)}, D_k^{(i)} + C_k^{(i)} P_{k|k}^{(j)} C_k^{(i)T}} (\zeta_k^{(i)}; m_{k|k}^{(j)})$$

$$\acute{B}_{k|l} = \sum_{j=1}^{J_{R,k|k-1}} \sum_{i=1}^{J_{k|l}} w_{R,k|k-1}^{(j)} w_k^{(i)} \times \mathcal{N}_{C_k^{(i)}, D_k^{(i)} + C_k^{(i)} P_{R,k|k-1}^{(j)} C_k^{(i)T}} (\zeta_k^{(i)}; m_{R,k|k-1}^{(j)})$$

$$\H{B}_{k|l}(z) = \sum_{j=1}^{J_{R,k|k-1}} \sum_{i=1}^{J_{k|l}} w_{R,k|k-1}^{(j)} w_k^{(i)} \times \mathcal{N}_{\H{C}_k^{(i)}, \H{D}_k^{(i,j)}} ([\zeta_k^{(i)T}, z^T]^T; m_{R,k|k-1}^{(j)})$$

$$\H{C}_k^{(i)} = \begin{bmatrix} C_k^{(i)} \\ H_k \end{bmatrix},$$

$$\H{D}_k^{(i,j)} = \begin{bmatrix} D_k^{(i)} & 0 \\ 0 & R_k \end{bmatrix} + \H{C}_k^{(i)} P_{R,k|k-1}^{(j)} \H{C}_k^{(i)T}$$

Moreover, the smoothed density is a Gaussian mixture given by (56)-(60).

F. Probability Hypothesis Density Smoothing

In a multi-target scenario the number of states and the states themselves vary with time in a random fashion. This is compounded by false measurements, detection uncertainty and data association uncertainty. The PHD (Probability Hypothesis Density) filter is a multi-target tracking solution that propagates the PHD (or intensity function) of the multi-target state forward in time [20], [28], [29], [21]. Recently, a *forward-backward PHD smoother* has been derived [25], [23].

The underlying model for the PHD filter includes all model parameters of the Bernoulli model except for the entry/reentry probability $p_{R,k|k-1}$ and entry/reentry state density

$f_{R,k|k-1}(\cdot)$. Instead the appearance of new targets is modelled by a random finite set of births parameterized by:

$$\gamma_{k|k-1} = \text{intensity of birth at time } k \tag{76}$$

At time k, the expected number of new targets is given by $\langle \gamma_{k|k-1}, 1 \rangle$ and the density of the new target state is given by the normalised $\gamma_{k|k-1}$, i.e. $\gamma_{k|k-1}/\langle \gamma_{k|k-1}, 1 \rangle$.

Similar to standard forward-backward smoothing, PHD smoothing consists of three steps: prediction update and backward smoothing. However, the actual PHD propagation equations are different to those for probability density propagations. For each integer $k > 0$, let $v_{k|l}$ denote the PHD at time k given the observation history $Z_{1:l}$ up to time l. Then, the prediction, update and backward smoothing steps are given respectively by:

$$v_{k|k-1}(\zeta) = \gamma_{k|k-1}(\zeta) + \langle v_{k-1|k-1} p_{S,k|k-1}, f_{k|k-1}(\zeta|\cdot) \rangle \tag{77}$$

$$v_{k|k}(x) = \left(q_{D,k}(x) + \sum_{z\in Z} \frac{p_{D,k}(x) g_k(z|x)}{\kappa_k(z) + \langle p_{D,k} g_k(z|\cdot), v_{k|k-1} \rangle} \right) \times v_{k|k-1}(x) \tag{78}$$

$$v_{k-1|l}(x) = \left(q_{S,k|k-1}(x) + p_{S,k|k-1}(x) \left\langle \frac{v_{k|l}}{v_{k|k-1}}, f_{k|k-1}(\cdot|x) \right\rangle \right) \times v_{k-1|k-1}(x) \tag{79}$$

A *linear Gaussian multi-target* (LG-MT) model assumes Gaussian mixture initial prior, constant probability of survival and probability of detection, linear Gaussian transition kernel and likelihood function for each target, as in the LG-Bernoulli model. Further, for the target birth model, the LG-MT model assumes a Gaussian mixture birth intensity:

$$\gamma_{k|k-1}(x) = \sum_{i=1}^{J_{\gamma,k|k-1}} w_{\gamma,k|k-1}^{(i)} \mathcal{N}(x; m_{\gamma,k|k-1}^{(i)}, P_{\gamma,k|k-1}^{(i)}), \tag{80}$$

where $J_{\gamma,k|k-1}$, $w_{\gamma,k|k-1}^{(i)}$, $m_{\gamma,k|k-1}^{(i)}$, $P_{\gamma,k|k-1}^{(i)}$, $i = 1, \ldots, J_{\gamma,k|k-1}$, are given model parameters that determine the shape of the birth intensity.

The PHD is intrinsically multi-modal, indeed, under LG-MT assumptions it was shown in [29] that if the initial PHD is a Gaussian mixture, then all subsequent predicted PHD and filtered PHD are Gaussian mixtures of the form:

$$v_{k|k-1}(x) = \sum_{i=1}^{J_{k|k-1}} \omega_{k|k-1}^{(i)} \mathcal{N}(x; \mu_{k|k-1}^{(i)}, \Pi_{k|k-1}^{(i)}), \tag{81}$$

$$v_{k|k}(x) = \sum_{i=1}^{J_{k|k}} \omega_{k|k}^{(i)} \mathcal{N}(x; \mu_{k|k}^{(i)}, \Pi_{k|k}^{(i)}), \tag{82}$$

G. The GMPHD smoother

This subsection presents the Gaussian mixture PHD smoothing solution for the linear Gaussian multi-target model described in subsection III-F. Here we deal with a PHD or intensity function rather than a probability density function.

Similar to the previous smoothing solutions, it is more convenient to rewrite the PHD update (78) and backward smoothing (79) in backward corrector form:

$$v_{k|k}(x) = v_{k|k-1}(x) L_k(Z_k; x) \tag{83}$$

$$v_{k|l}(x) = v_{k|k}(x) B_{k|l}(x) \tag{84}$$

where

$$L_k(Z; x) = q_{D,k}(x) + \sum_{z\in Z} \frac{p_{D,k}(x) g_k(z|x)}{\kappa_k(z) + \langle p_{D,k} g_k(z|\cdot), v_{k|k-1} \rangle} \tag{85}$$

$$B_{k|l}(x) = q_{S,k+1|k}(x) + p_{S,k+1|k}(x) \left\langle \frac{v_{k+1|l}}{v_{k+1|k}}, f_{k+1|k}(\cdot|x) \right\rangle \tag{86}$$

Note that the backward corrector starts with $B_{l|l}(x) = 1$.

The key to the closed form backward PHD recursion is the following recursion for the backward corrector term $B_{k|l}$

Proposition 10: For $k \le l$

$$B_{k-1|l}(x) = q_{S,k|k-1}(x) + p_{S,k|k-1}(x) \langle B_{k|l} L_k(Z_k; \cdot), f_{k|k-1}(\cdot|x) \rangle \tag{87}$$

Proof: It follows from (84) and (83) that

$$\frac{v_{k|l}}{v_{k|k-1}} = \frac{v_{k|k} B_{k|l}}{v_{k|k-1}} = L_k(Z_k; \cdot) B_{k|l},$$

which upon substitution into (86) with k replaced by $k-1$ gives (87). □

The forward PHD recursion (77), (83) and backward corrector recursion (87) can be thought of as some kind of "two-filter" PHD smoother. From Proposition 10, the recursion for the backward corrector falls under the generic form (41)-(42) with $p_k = p_{S,k|k-1}$ and $q_k = q_{S,k|k-1}$. Hence, from Proposition 6, we have the following result (for completeness the proof is given in Appendix VII-B).

Corollary 11: Under linear Gaussian multi-target assumptions, suppose that at time k, the PHD backward corrector has the form (43) then

$$B_{k-1|l}(x) = q_{S,k|k-1} + p_{S,k|k-1} \times \left(q_{D,k} \dot{B}_{k|l}(x) + p_{D,k} \sum_{z\in Z_k} \frac{\ddot{B}_{k|l}(x; z)}{\kappa_k(z) + p_{D,k} \eta_k(z)} \right) \tag{88}$$

where $\dot{B}_{k|l}(x)$, $\ddot{B}_{k|l}(x)$ are given respectively by (45), (48) and

$$\eta_k(z) = \sum_{j=1}^{J_{k|k-1}} \omega_{k|k-1}^{(j)} \mathcal{N}_{H_k, R_k + H_k \Pi_{k|k-1}^{(j)} H_k^T}(z; \mu_{k|k-1}^{(j)}) \tag{89}$$

Moreover, the smoothed PHD is a Gaussian mixture given by

$$p_{k|l}(x) = \sum_{i=1}^{J_{k|l}} \sum_{j=1}^{J_{k|k}} \omega_{k|k}^{(j)} q_{k|k}^{(i,j)}(\zeta_k^{(i)}) w_k^{(i)} \mathcal{N}(x; \tilde{\mu}_{k|k}^{(i,j)}(\zeta_k), \tilde{\Pi}_{k|k}^{(i,j)}), \tag{90}$$

where

$$q_{k|k}^{(i,j)}(\zeta_k^{(i)}) = \mathcal{N}_{C_k^{(i)}, D_k^{(i)} + C_k^{(i)} \Pi_{k|k}^{(j)} C_k^{(i)T}}(\zeta_k^{(i)}; \mu_{k|k}^{(j)}) \tag{91}$$

$$\tilde{\mu}_{k|k}^{(i,j)}(\zeta_k^{(i)}) = \mu_{k|k}^{(j)} + K_{k|k}^{(i,j)}(\zeta_k^{(i)} - C_k^{(i)} \mu_{k|k}^{(j)}) \tag{92}$$

$$\tilde{\Pi}_{k|k}^{(i,j)} = (I - K_{k|k}^{(i,j)} C_k^{(i)}) \Pi_{k|k}^{(j)} \tag{93}$$

$$K_{k|k}^{(i,j)} = \Pi_{k|k}^{(j)} C_k^{(i)T} (C_k^{(i)} \Pi_{k|k}^{(j)} C_k^{(i)T} + D_k^{(i)})^{-1} \tag{94}$$

IV. Canonical solution

The solutions presented so far are recursive, i.e. the backward corrector at a given time is calculated from the backward corrector at the previous time. It is of interest to obtain an explicit (canonical) expression for the backward corrector. Indeed, the canonical backward corrector is a mixture of Gaussians in some linear transformations of the state. Moreover, the matrices that parameterise each Gaussian component do not depend on the measurements and can be pre-computed analogous to the Kalman gain.

In this section we derive canonical expressions for the backward correctors described in the previous section. Subsection IV-A presents a set of terse yet suggestive notations that have natural interpretations in terms of measurement predictions which enable the derivation of the canonical expression for the backward corrector in subsection IV-B. Readers interested in numerical results only can skip this section.

A. Measurement prediction

The notion of measurement prediction described in this subsection facilitates the derivation of the specific formulae for the closed form canonical solutions. Given the measurement matrix H_i and measurement covariance R_i for $i > 0$, define

$$H_{i|i} \triangleq H_i, \quad R_{i|i} \triangleq R_i, \tag{95}$$

and for $i, j > 0$ such that $i \geq j$, define

$$H_{i|j-1} \triangleq H_{i|j} F_{j|j-1}, \tag{96}$$

$$R_{i|j-1} \triangleq R_{i|j} + H_{i|j} Q_j H_{i|j}^T. \tag{97}$$

We also use the obvious short hand $[H, R]_{i|j}$ when we refer to the pair $H_{i|j}, R_{i|j}$ collectively. Thus, to construct $[H, R]_{i|j}$, we start with $[H, R]_{i|i}$ from (95), then repeatedly applying (96), (97) to construct $[H, R]_{i|i-1}$, and $[H, R]_{i|i-2}$, and so forth, until we reach $[H, R]_{i|j}$. Note that $H_{i|j} = H_i F_{i|i-1} \cdots F_{j+1|j}$.

The matrices $H_{i|j}$ and $R_{i|j}$ defined by the recursion (96), (97) capture the statistics of the measurement at time i conditional on the state at time j, in the following sense (see Appendix VII-C for the proof).

Lemma 12: Under the linear Gaussian dynamic and measurement model (8), (9), given the state x_j at time j, the measurement z_i at time $i \geq j$ is Gaussian distributed with mean $H_{i|j} x_j$, and covariance $R_{i|j}$, i.e. $g_{i|j}(z_i|x_j) = \mathcal{N}_{H_{i|j}, R_{i|j}}(z_i; x_j)$.

Hence, the matrices $H_{i|j}$ and $R_{i|j}$ can be interpreted as the predicted measurement matrix and predicted measurement covariance to time i given the state at time j.

We now generalize the notion of measurement prediction to joint measurements. For $j > 0$ define

$$H_{\emptyset|j} = [], \quad R_{\emptyset|j} = []. \tag{98}$$

where [] is the MATLAB notation for the null matrix. Consider the set of integers $I = \{i(1), ..., i(|I|)\}$, where $|I|$ denotes the cardinality of I, and by convention $i(1) > i(2) > ... > i(|I|)$. In various places, we use the notation $k : l$ to denote the set of consecutive integers $\{k, ..., l\}$. Given I and $j > 0$ with $i(|I|) > j$ or $I = \emptyset$, define

$$H_{I\cup\{j\}|j} \triangleq \begin{bmatrix} H_{I|j} \\ H_j \end{bmatrix}, \quad R_{I\cup\{j\}|j} \triangleq \begin{bmatrix} R_{I|j} & 0 \\ 0 & R_j \end{bmatrix}, \tag{99}$$

and similarly to (96), (97)

$$H_{I|j-1} \triangleq H_{I|j} F_{j|j-1} \tag{100}$$

$$R_{I|j-1} \triangleq R_{I|j} + H_{I|j} Q_j H_{I|j}^T \tag{101}$$

Note that $H_{\{j\}|j} = H_j$, $R_{\{j\}|j} = R_j$, $H_{\{i\}|j} = H_{i|j}$, and $R_{\{i\}|j} = R_{i|j}$. Again, we use the obvious short hand notation $[H, R]_{I|j}$ when we refer to the pair $H_{I|j}, R_{I|j}$ collectively. Thus, to construct $[H, R]_{I|j}$ we start with $[H, R]_{\{i(1)\}|i(1)}$ and apply the *retrodiction* operation (100), (101) an appropriate number of times to obtain $[H, R]_{\{i(1)\}|i(2)}$; then using the *stacking* operation (99), to construct $[H, R]_{\{i(1),i(2)\}|i(2)}$ and apply (100), (101) an appropriate number of times to obtain $[H, R]_{\{i(1),i(2)\}|i(3)}$; and so forth.

We denote the joint measurements at times I by $z_I = [z_{i(1)}^T, ..., z_{i(|I|)}^T]^T$, i.e. the joint measurement space is $Z_I = Z_{i(1)} \times ... \times Z_{i(|I|)}$. The matrices $H_{I|j}, R_{I|j}$ defined by the recursions (99), and (100), (101) capture the statistics of the joint measurement z_I at times I, conditional on the state at time j, in the following sense (see Appendix VII-C for the proof).

Lemma 13: Under the linear Gaussian dynamic and measurement model (8), (9), given the state x_j at time j, the joint measurement z_I at times I, with $i(|I|) \geq j$, is Gaussian distributed with mean $H_{I|j} x_j$, and covariance $R_{I|j}$, i.e. $g_{I|j}(z_I|x_j) = \mathcal{N}_{H_{I|j}, R_{I|j}}(z_I; x_j)$.

Hence, the matrices $H_{I|j}$ and $R_{I|j}$ can be interpreted as the predicted measurement matrix and predicted measurement covariance to times I given the state at time j.

B. Canonical solution for the Generic Backward Corrector

In what follows we use the following notation for multiple sums:

$$\sum_{z_I \in Z_I} f(z_I) \triangleq \sum_{z_{i(1)} \in Z_{i(1)}} \cdots \sum_{z_{i(n)} \in Z_{i(|I|)}} f(z_{i(1)}, ..., z_{i(|I|)}),$$

with the convention $\sum_{z_\emptyset \in Z_\emptyset} f(z_\emptyset) = 1$ (this is not in conflict with $\sum_{z \in \emptyset} f(z) = 0$). The canonical form of the generic backward corrector is given by the following proposition (see Appendix VII-C for the proof).

Proposition 14: The closed form solution to the generic backward corrector recursion (41), (42) for $k \leq l$ is given by

$$B_{k|l}(x) = \sum_{I \subseteq \{l:k+1\}} \sum_{z_I \in Z_I} w_{I|k}^+ w_I(z_I) w_{I|k}^- \mathcal{N}_{[H,R]_{I|k}}(z_I; x) \tag{102}$$

where

$$w_{I|k}^+ = \prod_{i=l}^{\max(I,k)+1} p_i \alpha_i + \sum_{j=l}^{\max(I,k)+2} q_j \prod_{i=j-1}^{\max(I,k)+1} p_i \alpha_i + q_{\max(I,k)+1} \tag{103}$$

$$w_I(z_I) = \prod_{i \in I} p_i w_i(z_i) \tag{104}$$

$$w_{I|k}^- = \prod_{i \in \{l:k+1\} - (I \cup \{l:\max(I,k)+1\})} p_i \alpha_i \tag{105}$$

The canonical form (102) of the backward corrector is a mixture of Gaussians in $H_{I|k}x$, where $H_{I|k}$ is the predicted measurement matrix to times I given the state at time k. The matrices $H_{I|k}$, $R_{I|k}$ that parameterise each Gaussian component do not depend on the measurements and can be pre-computed from model parameters $F_{k|k-1}$, Q_k, H_k, and R_k.

The index I in the above proposition is an ordered (but not necessarily consecutive) set of integers taken from $\{l : k+1\}$. The term $w_I(z_I)$ defined from the set I, is dependent on the data z_I, but the term $w^+_{I|k}$ defined from the set $\{l : \max(I,k)+1\}$, and $w^-_{I|k}$ defined from the remaining indices, are independent of the data. These three sets of indices form a partition of the set $\{l : k+1\}$. When I is empty, the partition simply reduces to $\{l : k+1\}$.

For the special case $q_i = 0$ the summation in $w^+_{I|k}$ vanishes and $w^+_{I|k}$ and $w^-_{I|k}$ can be combined into one product over the indices $\{l : k+1\} - I$ and hence

$$w^+_{I|k} w_I(z_I) w^-_{I|k} = \prod_{i\in\{l:k+1\}-I} p_i\alpha_i \prod_{i\in I} p_i w_i(z_i)$$

The LG-FSO model falls under this special case with

$$p_i = 1,\ \alpha_i = \frac{q_{D,i}\kappa_i^{Z_i}}{\nu_i(Z_i)},\ w_i(z_i) = \frac{p_{D,i}\kappa_i^{Z_i-\{z_i\}}}{\nu_i(Z_i)},$$
$$\nu_i(Z_i) = q_{D,i}\kappa_i^{Z_i} + p_{D,i}\sum_{z\in Z_i}\kappa_i^{Z_i-\{z\}}\nu_i(z).$$

Consequently, the canonical form of the backward corrector is given by the following result.

Corollary 15: Under the LG-FSO model, the backward corrector $B_{k|l}$ for $k \le l$ is given by (102) with

$$w^+_{I|k} w_I(z_I) w^-_{I|k} = \frac{\prod_{j\in\{l:k+1\}-I} q_{D,j}\kappa_j^{Z_j} \prod_{i\in I} p_{D,i}\kappa_i^{Z_i-\{z_i\}}}{\prod_{i\in\{l:k+1\}} \nu_i(Z_i)}$$

Remark: The canonical form of the backward corrector for the linear Gaussian model with Gaussian mixture initial prior is the special case $p_{D,i} = 1$, $(q_{D,i} = 0)$, $\kappa_i = 0$, and $Z_i = \{z_i\}$, where every term in the the double sum vanishes except that with $I = \{l : k+1\}$ (because a product over an empty set of indices is 1 by convention):

$$B_{k|l}(x) = \frac{\mathcal{N}_{[H,R]_{\{l:k+1\}|k}}(z_{\{l:k+1\}}; x)}{\prod_{i\in\{l:k+1\}} \nu_i(z_i)}$$

This is consistent with the recursive result in Proposition 2.

For the LG-Bernoulli model the result is much more complex. Nonetheless, the LG-Bernoulli backward corrector is still a special case of Proposition 14.

Corollary 16: Under the LG-Bernoulli model (53), the backward corrector $B_{k|l}$ for $k \le l$ is given by (102) with

$$q_j = \alpha_{S,j|l},\ p_i = \frac{\beta_{S,i|l}}{\nu_{i|l}},\ \alpha_i = \frac{q_{D,i}\kappa_i^{Z_i}}{\nu_i(Z_i)},$$
$$w_i(z_i) = \frac{p_{D,i}\kappa_i^{Z_i-\{z_i\}}}{\nu_i(Z_i)}.$$

Note that this solution is different from the previous one in the sense that the parameters q_{k+1}, p_{k+1} are not defined purely in terms of predefined constants or model parameters, but are defined from $\alpha_{S,k+1|l}$ and $\beta_{S,k+1|l}$, which in turn, are computed from the smoothed existence probability, density and corrector in the previous iteration.

The canonical PHD backward corrector is also a special case of Proposition 14.

Corollary 17: Under the linear Gaussian multi-target model, the PHD backward corrector $B_{k|l}$ for $k < l$ is given by (102) with

$$q_i = q_{S,i|i-1},\ p_i = p_{S,i|i-1},\ \alpha_i = q_{D,i},$$
$$w_i(z_i) = \frac{p_{D,i}}{\kappa_i(z_i) + p_{D,i}\eta_i(z_i)}.$$

V. Numerical Examples

This section illustrates the proposed closed form forward-backward smoothing solutions via examples drawn from target tracking applications. These examples serve as verifications of our closed form solutions and are not intended as numerical studies of forward-backward smoothing algorithms. The recursive form of the backward corrector is used in the computation. Exact implementation of backward smoothing is exponential in memory requirement as in the forward filtering step. To manage the number of mixture components, pruning and merging of components is performed at each time step using a weight threshold of $T' = 10^{-5}$, a merging threshold of $U' = 4m$, and a maximum of $J_{max} = 100$ Gaussian posterior components (see [31] for the exact meaning of these parameters). In the calculation of the backwards corrector, the number of Gaussians is capped to $I_{max} = 10000$ terms in Demonstration 1 for a Bernoulli model and $I_{max} = 50000$ terms in Demonstration 2 for the PHD, with preference given to those components with larger weights.

To quantify the estimation error, an appropriate metric, known as the optimal sub-pattern assignment (OSPA) metric [26] is used for both examples since they involve uncertain and time-varying number of targets. Let $d^{(c)}(x,y) := \min(c, \|x-y\|)$ for $x, y \in \mathcal{X}$, and Π_k denote the set of permutations on $\{1,2,\ldots,k\}$ for any positive integer k. Then, for $X = \{x_1,\ldots,x_m\}$ and $Y = \{y_1,\ldots,y_n\}$, the OSPA metric $\bar{d}_p^{(c)}$ is defined as follows:

$$\bar{d}_p^{(c)}(X,\hat{X}) \triangleq \left(\frac{1}{n}\left(\min_{\pi\in\Pi_n}\sum_{i=1}^{m} d^{(c)}(x_i,\hat{x}_{\pi(i)})^p + c^p(n-m)\right)\right)^{\frac{1}{p}}$$

if $m \le n$; $\bar{d}_p^{(c)}(X,\hat{X}) \triangleq \bar{d}_p^{(c)}(\hat{X},X)$ if $m > n$; and $\bar{d}_p^{(c)}(X,\hat{X}) \triangleq 0$ if $m = n = 0$. In this work we use $p = 1$ and $c = 100m$ (meters).

A target state comprises target position and velocity, denoted by $x_k = [\ p_{x,k}, p_{y,k}, \dot{p}_{x,k}, \dot{p}_{y,k}\ ]^T$ at time k. Target generated observations are position only, denoted by $z_k = [\ z_{x,k}, z_{y,k}\ ]^T$ at time k. The transition density and likelihood function are linear Gaussian, i.e. (8), (9) with

$$F_{k|k-1} = \begin{bmatrix} I_2 & \Delta I_2 \\ 0_2 & I_2 \end{bmatrix},\quad Q_k = \sigma_\nu^2 \begin{bmatrix} \frac{\Delta^4}{4} I_2 & \frac{\Delta^3}{2} I_2 \\ \frac{\Delta^3}{2} I_2 & \Delta^2 I_2 \end{bmatrix}$$
$$H_k = \begin{bmatrix} I_2 & 0_2 \end{bmatrix},\qquad R_k = \sigma_\varepsilon^2 I_2$$

where $\Delta = 1s$ is the sampling period, $\sigma_\nu = 1m/s^2$, $\sigma_\varepsilon = 10m$, I_n and 0_n denote the $n \times n$ identity and zero matrices respectively. Targets are observed within a two dimensional surveillance region $[-1000, 1000]m \times [-1000, 1000]m$. The noisy measurements are additionally subjected to miss detections and false alarms. Detection uncertainty is modeled by a probability of detection, $p_{D,k} = 0.98$. Clutter is modelled by a spatial Poisson process with intensity function $\kappa_k(z) = \lambda_c V u(z)$, where $\lambda_c = 1.75 \times 10^{-6} m^{-2}$ is the average intensity, $V = 4 \times 10^6 m^2$ is the 'volume' of the surveillance region, and $u(\cdot)$ is a uniform probability density over the surveillance region. Note that in the calculation of estimation errors for the filters and smoothers under consideration, only the positions are used.

A. Demonstration 1

This demonstration involves a Bernoulli state space model with the following parameters. If the target is currently present, it survives to the next time step with probability of survival $p_{S,k|k-1} = 0.99$, or hence dies with probability $1 - p_{S,k|k-1} = 0.01$. If the target is currently absent, it enters/re-enters the scene with probability $p_{R,k|k-1} = 0.01$ and state vector distributed according to $f_{R,k|k-1} = \mathcal{N}(\cdot; m_R, P_R)$ where $m_R = [\ -400, 10, 400, -10\ ]^T$ and $P_R = \text{diag}([\ 100,\ 10,\ 100,\ 10\ ]^T)^2$, or hence remains absent with probability $1 - p_{R,k|k-1} = 0.99$. The target appear at times $k = 10s$, dies at $k = 80s$ and follows the path shown in Figure 1.

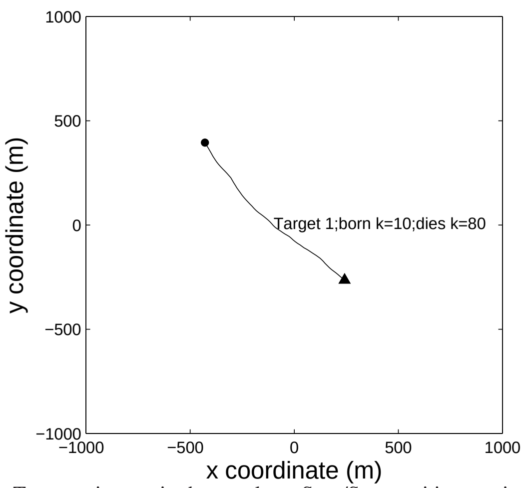

Fig. 1. Target trajectory in the xy plane. Start/Stop positions at times $k = 10$ and $k = 80$ are shown with ●/▲.

The initial prior is the Bernoulli random finite set density with zero existence probability and highly diffuse Gaussian state density. At each time, finite-set-valued state estimates are extracted according to the smoothed existence probability and state probability density. If the estimated probability of existence is less than 50%, an empty set is declared as an estimate. Otherwise, a singleton target set state estimate is declared as the mean of the Gaussian component with the highest filtered or smoothed weight respectively.

The Bernoulli filter and smoothers for lags of up to 3 steps are compared. Figure 2 shows the average OSPA errors over 1000 trials. These results confirm the observations that the filter initiates and terminates the track with a one step delay but otherwise performs well, as seen by the average OSPA error which peaks markedly at times $k = 10s$ and $k = 81s$ but is otherwise flat. The relatively flat average smoothing errors also confirm that each of the smoothers generally initiate and terminate the track at the correct times and again as expected produce more accurate state estimates with longer smoothing lags.

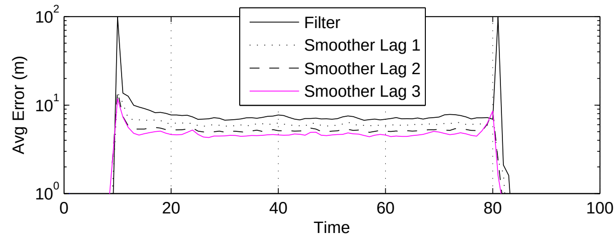

Fig. 2. Average OSPA errors for the forward filter and backward smoother with lags of 1,2 and 3 time steps.

B. Demonstration 2

This demonstration presents a multiple target tracking scenario where target can appear or disappear. The PHD forward filter and backward smoother are used to estimate the intensity function of the multiple target state. The parameters of the system model is the same as the previous example, with $p_{R,k|k-1}$ and $f_{R,k|k-1}$ replaced by a Poisson target births with Gaussian intensity $\gamma_k(x) = 0.04\mathcal{N}(x; 0, 100I_4)$.

The initial prior is the zero intensity function $v_0 = 0$. The number of targets is estimated by rounding the volume of the intensity function to the nearest integer. Multiple target state estimates are generated from the estimated intensity function by extracting the corresponding number of the means from the filtered or smoothed Gaussian components with the highest weights [29].

This scenario involves 4 targets for the entire duration. Each target starts at the origin and respectively heads north, south, east and west along the principal axes with a constant speed of $5ms^{-1}$ as indicated in Figure 3. The filter (and smoother) does not have a priori knowledge of the fixed number of targets.

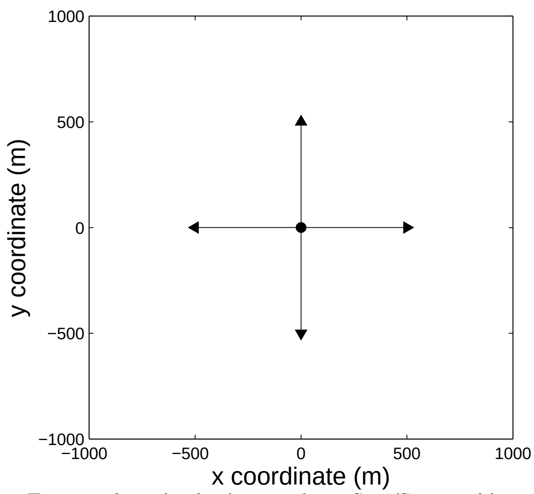

Fig. 3. Target trajectories in the xy plane. Start/Stop positions are shown with ●/▲.

Figure 4 shows the and OSPA error over 1000 Monte Carlo runs. It can be seen that the filter initially incurs some error in estimating the number of targets and their locations due to being initialized with a zero intensity. In the case of the filter, there is a small settling in period before the error stabilizes to a value indicating that the target tracks have been properly established. In the case of the smoothers compared to the filter, the settling in period is shorter and during this time the error incurred is lower. For the entire duration, the errors also decrease from the filter to the smoother and with increasing smoother lag. All of these results are consistent with expectations.

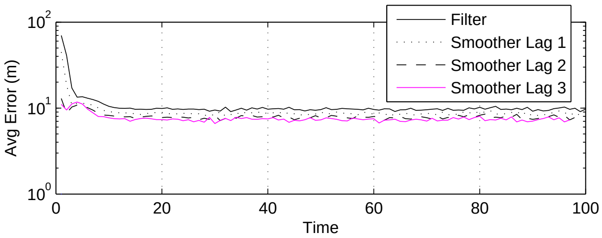

Fig. 4. Average OSPA errors for the forward filter and backward smoother with lags of 1,2 and 3 time steps.

VI. Conclusions

A Gaussian sum smoother and, more importantly, closed form smoothing solutions for: linear Gaussian model with finite set observations; linear Gaussian Bernoulli model; and the PHD under linear Gaussian multi-target assumptions, have been derived. These solutions are based on alternative forms of the forward-backward recursions in which the smoothed densities are the product of corresponding filtered densities and backward correctors. The backward correctors are mixtures of Gaussians in some linear transformations of the state and the smoothed densities are Gaussian mixtures. The matrices that parameterise each Gaussian component do not depend on the measurements and can be pre-computed analogous to the Gaussian sum filter. Numerical results have also been presented to verify our proposed closed form solutions.

VII. Appendix

A. Some standard Gaussian identities

Lemma 18: Given F, Q and H, R of appropriate dimensions, and that Q and R are positive definite

$$\begin{aligned}\langle \mathcal{N}_{H,R}(z;\cdot), \mathcal{N}(\cdot;Fx,Q)\rangle &= \langle \mathcal{N}_{H,R}(z;\cdot), \mathcal{N}_{F,Q}(\cdot;x)\rangle \\ &= \mathcal{N}_{HF,R+HQH^T}(z;x) \end{aligned} \quad (106)$$

Lemma 19: Given H, R, m, and P of appropriate dimensions, and that R and P are positive definite,

$$\mathcal{N}_{H,R}(z;x)\mathcal{N}(x;m,P) = \mathcal{N}(x;\tilde{m},\tilde{P})\mathcal{N}_{H,R+HPH^T}(z;m) \quad (107)$$

where

$$\tilde{m} = \tilde{m}(z,m) = m + K(z - Hm) \quad (108)$$

$$\tilde{P} = (I - KH)P \quad (109)$$

$$K = PH^T(HPH^T + R)^{-1} \quad (110)$$

B. Recursive solution

Proof of Proposition 6: Using the generic pseudo likelihood (42), we have

$$\begin{aligned} B_{k|l}(x)L_k(Z_k;x) &= \sum_{i=1}^{J_{k|l}} \alpha_k w_k^{(i)} \mathcal{N}_{C_k^{(i)},D_k^{(i)}}(\zeta_k^{(i)};x) \\ &+ \sum_{i=1}^{J_{k|l}} \sum_{z\in Z_k} w_k(z) w_k^{(i)} \mathcal{N}_{C_k^{(i)},D_k^{(i)}}(\zeta_k^{(i)};x)\mathcal{N}_{H_k,R_k}(z;x) \\ &= \sum_{i=1}^{J_{k|l}} \alpha_k w_k^{(i)} \mathcal{N}_{C_k^{(i)},D_k^{(i)}}(\zeta_k^{(i)};x) \\ &+ \sum_{i=1}^{J_{k|l}} \sum_{z\in Z_k} w_k(z) w_k^{(i)} \mathcal{N}_{\tilde{C}_k^{(i)},\tilde{D}_k^{(i)}}([\zeta_k^{(i)T}, z^T]^T;x) \end{aligned} \quad (111)$$

where

$$\tilde{C}_k^{(i)} = \begin{bmatrix} C_k^{(i)} \\ H_k \end{bmatrix}, \quad \tilde{D}_k^{(i)} = \begin{bmatrix} D_k^{(i)} & 0 \\ 0 & R_k \end{bmatrix}$$

Hence, using the generic backward corrector recursion (41),

$$\begin{aligned} B_{k-1|l}(x) = \; & q_k + \sum_{i=1}^{J_{k|l}} p_k \alpha_k w_k^{(i)} \left\langle \mathcal{N}_{C_k^{(i)},D_k^{(i)}}(\zeta_k^{(i)};\cdot), \mathcal{N}_{F_{k|k-1},Q_k}(\cdot;x)\right\rangle + \\ & \sum_{i=1}^{J_{k|l}} \sum_{z\in Z_k} p_k w_k(z) w_k^{(i)} \left\langle \mathcal{N}_{\tilde{C}_k^{(i)},\tilde{D}_k^{(i)}}([\zeta_k^{(i)T}, z^T]^T;\cdot), \mathcal{N}_{F_{k|k-1},Q_k}(\cdot;x)\right\rangle \end{aligned}$$

and applying Lemma 18 completes the proof. □

Proof of Corollary 7: Under the LG-FSO model, the predicted density is a Gaussian mixture of the form (13). Moreover, using (53),

$$\begin{aligned} \langle g_k(Z;\cdot), p_{k|k-1}\rangle &= q_{D,k}\kappa_k^Z \\ &+ p_{D,k} \sum_{z\in Z} \kappa_k^{Z-\{z\}} \langle \mathcal{N}_{H_k,R_k}(z;\cdot), p_{k|k-1}\rangle \\ &= q_{D,k}\kappa_k^Z + p_{D,k} \sum_{z\in Z} \kappa_k^{Z-\{z\}} \nu_k(z) \end{aligned}$$

where the last equality follows from (25). Hence,

$$L_k(Z_k;x) = \frac{q_{D,k}\kappa_k^{Z_k} + p_{D,k} \sum_{z\in Z_k} \kappa_k^{Z_k-\{z\}} \mathcal{N}_{H_k,R_k}(z;x)}{q_{D,k}\kappa_k^{Z_k} + p_{D,k} \sum_{z\in Z_k} \kappa_k^{Z_k-\{z\}} \nu_k(z)}. \quad (112)$$

Consequently, the result follows from Proposition 6 with $q_k = 0$, $p_k = 1$. The expression for the smoothed density follows from along the same line as Corollary 3. □

Proof of Corollary 9: Under the linear Gaussian Bernoulli model, the filtered density is Gaussian mixture of the form (14), and hence,

$$\begin{aligned} &\langle p_{k|k}, B_{k|l}\rangle \\ &= \sum_{j=1}^{J_{k|k}} \sum_{i=1}^{J_{k|l}} w_{k|k}^{(j)} w_k^{(i)} \left\langle \mathcal{N}_{C_k^{(i)},D_k^{(i)}}(\zeta_k^{(i)};\cdot), \mathcal{N}(\cdot;m_{k|k}^{(j)}, P_{k|k}^{(j)})\right\rangle \\ &= \nu_{k|l} \end{aligned}$$

where the last equality follows from Lemma 18.

The pseudo-likelihood is the same as (112). Hence, it follows from Proposition 8 that the recursion for the backward corrector falls under the generic form of the Gaussian mixture smoother with $q_k = \alpha_{S,k|l}$ and $p_k = \beta_{S,k|l}/\nu_{k|l}$. Consequently, the expression for $B_{k-1|l}$ follows from Proposition 6. The expression for the smoothed density follows from along the same line as Corollary 3.

For the probability of existence $r_{k-1|l}$, recall from the proof of Proposition 6, the expression for $B_{k|l}(x)L_k(Z_k;x)$ in (111). Substitute (111) and (71) into (75) and applying Lemma 18 completes the proof. □

Proof of Corollary 11: Under LG-MT model assumptions, the predicted PHD is a Gaussian mixture of the form (13). Hence

$$\begin{aligned}&\langle \mathcal{N}_{H_k,R_k}(z;\cdot), v_{k|k-1}\rangle\\&= \sum_{j=1}^{J_{k|k-1}} \omega_{k|k-1}^{(j)} \left\langle \mathcal{N}_{H_k,R_k}(z;\cdot), \mathcal{N}(\cdot;\mu_{k|k-1}^{(j)}, \Pi_{k|k-1}^{(j)})\right\rangle\\&= \eta_k(z)\end{aligned}$$

by virtue of Lemma 18. Hence, the pseudo likelihood is

$$\begin{aligned}L_k(Z;x) &= q_{D,k} + \sum_{z\in Z} \frac{p_{D,k}\mathcal{N}_{H_k,R_k}(z;x)}{\kappa_k(z) + p_{D,k}\langle \mathcal{N}_{H_k,R_k}(z;\cdot), v_{k|k-1}\rangle}\\&= q_{D,k} + \sum_{z\in Z} \frac{p_{D,k}\mathcal{N}_{H_k,R_k}(z;x)}{\kappa_k(z) + p_{D,k}\eta_k(z)}\end{aligned}$$

From Proposition 10, the recursion for the backward corrector falls under the generic form of the GM smoother with $p_k = p_{S,k|k-1}$ and $q_k = q_{S,k|k-1}$. Hence the result follows from Proposition 6. The expression for the smoothed density follows from along the same line as Corollary 3. □

C. Canonical solution

Proof of Lemma 12: First note from the Markov assumption that the joint density of $x_{i:j}, z_{i:j-1}$ conditional on x_{j-1} is given by

$$\begin{aligned}&p(x_{i:j}, z_{i:j-1}|x_{j-1})\\&= \prod_{k=i-1}^{j-1} f_{k+1|k}(x_{k+1}|x_k) \prod_{k=i}^{j-1} g_k(z_k|x_k)\\&= p(x_{i:j+1}, z_{i:j}|x_j) f_{j|j-1}(x_j|x_{j-1}) g_{j-1}(z_{j-1}|x_{j-1}) \qquad (113)\end{aligned}$$

and hence

$$\begin{aligned}&g_{i|j-1}(z_i|x_{j-1})\\&= \iint p(x_{i:j}, z_{i:j-1}|x_{j-1}) dz_{i-1:j-1} dx_{i:j}\\&= \iint p(x_{i:j+1}, z_{i:j}|x_j) f_{j|j-1}(x_j|x_{j-1}) g_{j-1}(z_{j-1}|x_{j-1}) dz_{i-1:j-1} dx_{i:j}\\&= \int \left[\iint p(x_{i:j+1}, z_{i:j}|x_j) dz_{i-1:j} dx_{i:j+1}\right] f_{j|j-1}(x_j|x_{j-1}) dx_j\\&= \int g_{i|j}(z_i|x_j) f_{j|j-1}(x_j|x_{j-1}) dx_j\end{aligned}$$

Given any i, suppose that the result holds i.e. for $j \le i$, $g_{i|j}(z_i|x_j) = \mathcal{N}_{H_{i|j},R_{i|j}}(z_i;x_j)$, then it follows from the linear Gaussian transition kernel (8) and the above equation that

$$\begin{aligned}g_{i|j-1}(z_i|x_{j-1}) &= \langle \mathcal{N}_{H_j,R_j}(z_j;\cdot), \mathcal{N}_{F_{j|j-1}Q_j}(\cdot;x_{j-1})\rangle\\&= \mathcal{N}_{H_{i|j-1},R_{i|j-1}}(z_i;x_{j-1})\end{aligned}$$

by Lemma 18. Thus the result also holds for $j-1$ Moreover, by virtue of the linear Gaussian measurement likelihood (9) the result is true for $j = i$, hence the lemma holds by induction. □

Proof of Lemma 13: Note that the stacking and retrodiction operations (99), and (100), (101) and can be used to generate $[H,R]_{I|j}$ for any index set I (whose elements are arranged in descending order) and j with $i\,(|I|) > j$. Hence, it suffices to show that

$$\begin{aligned}g_{I|j}(z_I|x_j) &= \mathcal{N}_{[H,R]_{I|j}}(z_I;x_j) \Rightarrow\\g_{I\cup\{j\}|j}(z_{I\cup\{j\}}|x_j) &= \mathcal{N}_{[H,R]_{I\cup\{j\}|j}}(z_{I\cup\{j\}|j};x_j) \qquad (114)\\g_{I|j}(z_I|x_j) &= \mathcal{N}_{[H,R]_{I|j}}(z_I;x_j) \Rightarrow\\g_{I|j-1}(z_I|x_{j-1}) &= \mathcal{N}_{[H,R]_{I|j-1}}(z_{I|j-1};x_{j-1}) \qquad (115)\end{aligned}$$

Given I and j with $i(|I|) > j$, (115) can be shown using similar arguments as in the proof of Lemma 12. It remains to show (114)

To show (114), note that $g_{I|j}(z_I|x_j)$ and $g_{I\cup\{j\}|j}(z_{I\cup\{j\}}|x_j)$ can be obtained by marginalising $p(x_{i(1):j+1}, z_{i(1):j}|x_j)$ over $x_{i(1):j+1}$ and the z_i whose indices are not in I and $I\cup\{j\}$ respectively. Using (113) we have

$$\begin{aligned}&g_{I\cup\{j\}|j}(z_{I\cup\{j\}}|x_j)\\&= \int \prod_{k=i(1)-1}^{j} f_{k+1|k}(x_{k+1}|x_k) \prod_{k\in I\cup\{j\}} g_k(z_k|x_k) dx_{i(1):j+1}\\&= \int \prod_{k=i(1)-1}^{j} f_{k+1|k}(x_{k+1}|x_k) \prod_{k\in I} g_k(z_k|x_k) g_j(z_j|x_j) dx_{i(1):j+1}\\&= g_{I|j}(z_I|x_j) g_j(z_j|x_j)\end{aligned}$$

Thus it follows from the linear Gaussian likelihood function (9) that

$$\begin{aligned}g_{I\cup\{j\}|j}(z_{I\cup\{j\}}|x_j) &= \mathcal{N}_{H_{I|j},R_{I|j}}(z_I;x_j)\mathcal{N}_{H_j,R_j}(z_j;x_j)\\&= \mathcal{N}_{[H,R]_{I\cup\{j\}|j}}(z_{I\cup\{j\}|j};x_j). \square\end{aligned}$$

Proof of Proposition 14: For convenience, let $w_{I|k}(z_I) = w^+_{I|k} w_I(z_I) w^-_{I|k}$. The result holds for the initial step $k = l$, since applying Proposition 6 with $B_{l|l} = 1$, gives

$$\begin{aligned}B_{l-1|l}(x) &= q_l + p_l\alpha_l + \sum_{z_l\in Z_l} p_l w_l(z_l)\mathcal{N}_{[H,R]_{l|l-1}}(z_l;x)\\&= \sum_{I\subseteq\{l\}} \sum_{z_I\in Z_I} w_{I|l-1}(z_I)\mathcal{N}_{[H,R]_{I|l-1}}(z_I;x) \quad (116)\end{aligned}$$

where

$$w_{I|l-1}(z_I) = \begin{cases} p_l w_l(z_l), I = \{l\}\\ (q_l + p_l\alpha_l), I = \emptyset\end{cases} \qquad (117)$$

Suppose that the result hold for $k+1$, i.e.

$$B_{k+1|l}(x) = \sum_{J\subseteq\{l:k+2\}} \sum_{z_J\in Z_J} w_{J|k+1}(z_J)\mathcal{N}_{[H,R]_{J|k+1}}(z_J;x)$$

where for $J \subseteq \{l : k+2\}$

$w_{J|k+1}(z_J) =$

$$\left(\prod_{i=l}^{\max(J,k+1)+1} p_i\alpha_i + \sum_{j=l}^{\max(J,k+1)+2} q_j \prod_{i=j-1}^{\max(J)+1} p_i\alpha_i + q_{\max(J,k+1)+1}\right) \times \left(\prod_{i\in I} p_i w_i(z_i)\right)\left(\prod_{i\in\{l:k+2\}-(J\cup\{l:\max(J,k+1)+1\})} p_i\alpha_i\right)$$

Then from Proposition 6, and the definitions (100), (101), (99), $B_{k|l}$ is given by

$$B_{k|l}(x) = q_{k+1} + p_{k+1}\alpha_{k+1}\sum_{J\subseteq\{l:k+2\}}\sum_{z_J\in Z_J} w_{J|k+1}(z_J)\mathcal{N}_{[H,R]_{J|k+1}}(z_J;x) + p_{k+1}\sum_{z\in Z_{k+1}} w_{k+1}(z)\sum_{J\subseteq\{l:k+2\}}\sum_{z_J\in Z_J} w_{J|k+1}(z_J)\mathcal{N}_{[H,R]_{J\cup\{k+1\}|k+1}}([z_J^T, z^T]^T; x)$$

which can be expressed in the form (102) by setting

$$w_{I|k}(z_I) = \begin{cases} q_{k+1} + p_{k+1}\alpha_{k+1}w_{\emptyset|k+1}(z_\emptyset), I = J = \emptyset \\ p_{k+1}\alpha_{k+1}w_{J|k+1}(z_I), I = J, I \neq \emptyset, J \subseteq \{l:k+2\} \\ p_{k+1}w_{k+1}(z_{k+1})w_{J|k+1}(z_J), I = J\cup\{k+1\}, J \subseteq \{l:k+2\} \end{cases} \tag{118}$$

for $I \subseteq \{l : k+1\}$. It remains to show that (118) is identical to (103).

For the case $I = J = \emptyset$, $\max(J, k+1) = k+1$ and

$$w_{\emptyset|k+1}(z_\emptyset) = \left(\prod_{i=l}^{k+2} p_i\alpha_i + \sum_{j=l}^{k+3} q_j \prod_{i=j-1}^{k+2} p_i\alpha_i + q_{k+2}\right),$$

$$w_{I|k}(z_I) = q_{k+1} + p_{k+1}\alpha_{k+1}w_{\emptyset|k+1}(z_\emptyset) = \prod_{i=l}^{k+1} p_i\alpha_i + \sum_{j=l}^{k+2} q_j \prod_{i=j-1}^{k+1} p_i\alpha_i$$

For the case $I = J, I \neq \emptyset, J \subseteq \{l : k+2\}$, note that $\max(I, k) = \max(I) = \max(J) = \max(J, k+1)$. Hence

$$w_{J|k+1}(z_J) = \left(\prod_{i=l}^{\max(I,k)+1} p_i\alpha_i + \sum_{j=l}^{\max(I,k)+1} q_j \prod_{i=j-1}^{\max(I,k)+1} p_i\alpha_i\right) \times \left(\prod_{i\in J} p_i w_i(z_i)\right)\left(\prod_{i\in\{l:k+2\}-(J\cup\{l:\max(I,k)+1\})} p_i\alpha_i\right) \tag{119}$$

Consequently,

$$\begin{aligned} w_{I|k}(z_I) &= p_{k+1}\alpha_{k+1}w_{J|k+1}(z_J) \\ &= \left(\prod_{i=l}^{\max(I,k)+1} p_i\alpha_i + \sum_{j=l}^{\max(I,k)+1} q_j \prod_{i=j-1}^{\max(I,k)+1} p_i\alpha_i\right) \times \left(\prod_{i\in J} p_i w_i(z_i)\right)\left(\prod_{i\in\{l:k+1\}-(J\cup\{l:\max(I,k)+1\})} p_i\alpha_i\right) \\ &= \left(\prod_{i=l}^{\max(I,k)+1} p_i\alpha_i + \sum_{j=l}^{\max(I,k)+1} q_j \prod_{i=j-1}^{\max(I,k)+1} p_i\alpha_i\right) \times \left(\prod_{i\in I} p_i w_i(z_i)\right)\left(\prod_{i\in\{l:k+1\}-(I\cup\{l:\max(I,k)+1\})} p_i\alpha_i\right) \end{aligned}$$

For the third case $I = J \cup \{k+1\}, J \subseteq \{l : k+2\}$, note that $\max(J, k+1) = \max(I) = \max(I, k)$ unless J is empty (in which case the result holds trivially). Hence (119) holds. Moreover $\{l : k+2\} - (J \cup \{l : \max(I, k) + 1\}) = \{l : k+1\} - (I \cup \{l : \max(I, k) + 1\})$. Consequently

$$\begin{aligned} w_{I|k}(z_I) &= p_{k+1}w_{k+1}(z_{k+1})w_{J|k+1}(z_J) \\ &= \left(\prod_{i=l}^{\max(I,k)+1} p_i\alpha_i + \sum_{j=l}^{\max(I,k)+1} q_j \prod_{i=j-1}^{\max(I,k)+1} p_i\alpha_i\right) \times \left(\prod_{i\in J\cup\{k+1\}} p_i w_i(z_i)\right)\left(\prod_{i\in\{l:k+2\}-(J\cup\{l:\max(I,k)+1\})} p_i\alpha_i\right) \\ &= \left(\prod_{i=l}^{\max(I,k)+1} p_i\alpha_i + \sum_{j=l}^{\max(I,k)+1} q_j \prod_{i=j-1}^{\max(I,k)+1} p_i\alpha_i\right) \times \left(\prod_{i\in I} p_i w_i(z_i)\right)\left(\prod_{i\in\{l:k+1\}-(I\cup\{l:\max(I,k)+1\})} p_i\alpha_i\right). \end{aligned}$$

Therefore the result follows by induction. □

REFERENCES

[1] D. L. Alspach and H. W. Sorenson, "Nonlinear Bayesian estimation using Gaussian sum approximations," *IEEE Trans. AC*, Vol. AC-17, no. 4, pp. 439–448, 1972.

[2] B. O. Anderson, and J. B. Moore, *Optimal Filtering*, Prentice-Hall, New Jersey, 1979.

[3] Y. Bar-Shalom, X. Li, and T. Kirubarajan, *Estimation with Applications to Tracking and Navigation*, New York: Wiley, 2001.

[4] Y. Bresler, "Two- filter formula for discrete-time non-linear Bayesian smoothing," *International Journal of Control*, Vol. 43, No. 2, 629–641, 1986.

[5] M. Briers, A. Doucet, and S. Maskell, "Smoothing algorithms for state-space models," *Annals of the Institute of Statistical Mathematics*, Vol. 62, No. 1, pp. 61–89, Springer Netherlands, 2010.

[6] A. Doucet, S. Godsill, and C. Andrieu, "On sequential Monte Carlo sampling methods for Bayesian filtering," *Statistics and Computing*, Vol. 10, No. 3, pp. 197–208, 2000.

[7] A. Doucet, N. de Freitas, and N. Gordon. *Sequential Monte Carlo Methods in Practice*. Springer-Verlag, 2001.

[8] P. Fearnhead, D. Wyncoll, J. and Tawn, "A sequential smoothing algorithm with linear computational cost," *Technical report*, Department of Mathematics and Statistics, Lancaster University, 2008.

[9] W. Fong, S. J. Godsill, A. Doucet, and M. West, "Monte Carlo smoothing with application to audio signal enhancement," *IEEE Trans. on Signal Processing*, Vol. 50, No. 2, pp. 438–449, Feb. 2002.

[10] S. J. Godsill, A. Doucet, and M. West, "Monte Carlo smoothing for nonlinear time series," *J. the American Statistical Association*, Vol. 99, pp. 156–168, Mar. 2004.

[11] N. Gordon, D. Salmond, and Smith, "Novel approach to nonlinear/non-Gaussian Bayesian state estimation," *Proc. IEE*, Vol. 140, No. 2, pp. 107-113, 1993.

[12] A. C. Harvey, *Forecasting, Structural Time Series Models and the Kalman Filter*, New York: Cambridge University Press, 1989.

[13] M. Hürzeler, and H. R. Künsch, "Monte Carlo approximations for general state space models," *Journal of Computational and Graphical Statistics*, Vol. 7, No. 2, pp. 175–193, Jun. 1998.

[14] R. Kalman "A new approach to linear filtering and prediction problems," *Transactions of the ASME–Journal of Basic Engineering*, Vol. 82, no. Series D, pp. 35–45, 1960.

[15] G. Kitagawa, "Non-Gaussian state space modelling of nonstationary time series," *Journal of American Statistical Association*, Vol. 82, pp. 1032-1063, 1987.

[16] G. Kitagawa,"The two-filter formula for smoothing and an implementation of the Gaussian sum smoother", *Ann. Inst. Statist. Math.*, Vol. 46, No. 4, pp. 605-623, 1994

[17] G. Kitagawa, "Monte Carlo filter and smoother for nonlinear non-Gaussian state-space models," *Journal of Computational Graphical Statistics*, Vol. 5, No. 1, pp. 1–25, Mar. 1996.

[18] G. Kitagawa, and G. Gersch, "Smoothness priors analysis of time series", *Lecture notes in Statistics*, 116 Springer, 1996.

[19] J. T.-H. Lo, "Finite-dimensional sensor orbits and optimal non-linear filtering," *IEEE Trans. IT*, vol. IT-18, no. 5, pp. 583–588, 1972.

[20] R. Mahler, "Multi-target Bayes Filtering via First-Order Multi-target Moments", *IEEE Trans. Aerospace & Electronic Systems*, Vol. 39, No. 4, pp. 1152-1178, 2003.

[21] R. Mahler, *Statistical Multisource Multitarget Information Fusion*, Artech House, 2007.

[22] R. Mahler, "PHD filters of higher order in target number," *IEEE Trans. Aerospace & Electronic Systems*, Vol. 43, No. 3, July 2007.

[23] R. Mahler, B.-N. Vo, and B.-T. Vo, "Multi-target forward-backward smoothing with the Probability Hypothesis Density," *IEEE Trans. Aerospace & Electronic Systems*, Vol. 48, No. 1, pp. 707-728, 2012.

[24] J. Meditch, "A survey of data smoothing for linear and nonlinear dynamic systems," *Automatica,* Vol. 9, 151–162, 1973.

[25] N. Nandakumaran, T. Kirubarajan, T. Lang, M. McDonald, and K. Punithakumar, "Multitarget tracking using probability hypothesis density smoothing", *IEEE Trans. Aerospace & Electronic Systems*, (to appear).

[26] D. Schuhmacher, B.-T. Vo, and B.-N. Vo, "A consistent metric for performance evaluation of multi-object filters," *IEEE Trans. Signal Processing*, Vol. 56, No. 8, pp. 3447–3457, Aug. 2008.

[27] H. W. Sorenson and D. L. Alspach, "Recursive Bayesian estimation using Gaussian sum," *Automatica*, Vol. 7, pp. 465–479, 1971.

[28] B.-N. Vo, S. Singh and A. Doucet, "Sequential Monte Carlo methods for Multi-target filtering with Random Finite Sets", *IEEE Trans. Aerospace & Electronic Systems*, Vol. 41, No. 4, pp. 1224–1245, 2005.

[29] B.-N. Vo, and W.-K. Ma, "The Gaussian mixture Probability Hypothesis Density filter," *IEEE Trans. Signal Processing*, Vol. 54, No. 11, pp. 4091-4104, 2006.

[30] B.-T. Vo, B.-N. Vo, and A. Cantoni, "Analytic implementations of the Cardinalized Probability Hypothesis Density filter," *EEE Trans. Signal Processing*, Vol. 55, No. 7, pp. 3553–3567, Jul. 2007.

[31] B.-T. Vo, B.-N Vo and A. Cantoni, "Bayesian Filtering with Random Finite Set Observations," *IEEE Trans. Signal Processing*, Vol. 56, No. 4, pp. 1313-1326, 2008.

[32] B.-T. Vo, *Random Finite Sets in Multi-Object Filtering,* PhD Thesis, University of Western Australia, 2008.

[33] B.-N. Vo, B.-T. Vo, and R. Mahler, "A Closed Form Solution to the Probability Hypothesis Density Smoother," *Proc. 13th Annual Conf. Information Fusion*, Edinburgh, UK, 2010.

[34] B.-T. Vo, C.-M. See, N. Ma, and W.-T. Ng, "Multi-sensor joint detection and tracking with the Bernoulli filter," *IEEE Trans. Aerospace & Electronic Systems*, Vol. 48, No. 2, 2012.

[35] B.-T. Vo, D. Clark, B.-N. Vo, and B. Ristic "Bernoulli Forward-Backward Smoothing for Joint Target Detection and Tracking," *IEEE Trans. Signal Processing*, Vol. 59, No. 9, pp. 4473-4477, 2011.

[36] M. West, and J. Harrison, *Bayesian forecasting and dynamic models*, Springer Series in Statistics, Springer New York, 1989.