# Extended Target Poisson Multi-Bernoulli Filter

Yuxuan Xia, Karl Granström, Member, IEEE, Lennart Svensson, Senior Member, IEEE, and Maryam Fatemi

Abstract—This paper presents a Poisson multi-Bernoulli (PMB) filter for multiple extended targets estimation. The extended target PMB filter is based on the Poisson multi-Bernoulli mixture (PMBM) conjugate prior for multiple extended target filtering and approximates the multi-Bernoulli (MB) mixture in the posterior density as a single MB. Because both the prediction and the update preserve the PMB form of the density, the proposed PMB filter is computationally cheaper than the PMBM filter while maintaining good filtering performance. Different methods for merging the MB mixture as a single MB are presented, along with their Gamma Gaussian inverse Wishart implementations. The performance of the extended target PMB filter is compared to the extended target PMBM filter and the extended target labelled MB filter in a thorough simulation study.

Index Terms—Multi-target tracking, Bayesian estimation, extended target, random finite sets, Kullback-Leibler divergence, random matrix model

#### I. INTRODUCTION

Multi-target tracking (MTT) denotes the process of estimating the set of target trajectories based on a sequence of noise-corrupted measurements [1]. Conventional MTT algorithms are usually tailored to the point target assumption: each target is modeled as a point without spatial extent, and each target gives rise to at most one measurement per time step. However, modern high-resolution radar and lidar sensors make the point target assumption unrealistic because with such sensors a target may give rise to multiple measurements per time step. The tracking of such a target leads to the extended target tracking problem, where the objective is to recursively estimate the target extent and kinematic states over time.

Random Finite Sets (RFS) [2] is a popular and widely used framework that facilitates an elegant Bayesian formulation of the MTT problem. In recent years, a significant trend in RFS-based MTT is the development of multi-target conjugate priors<sup>1</sup> that can be described by a finite set of parameters. The Poisson Multi-Bernoulli Mixture (PMBM) is a popular multi-target conjugate prior for both point targets and extended targets. In several simulation studies it has been shown that, compared to other RFS-based filters, the PMBM filters have state-of-the-art performance for tracking the set of targets, see, e.g., [4]–[7]. It is therefore well-motivated to develop tracking algorithms based on the PMBM conjugate prior.

The partitioning of noisy sensor measurements into potential tracks and false alarms, also known as data association, is an

inherent challenging in MTT. Due to the unknown number of data associations, the number of multi-Bernoulli (MB) components in the posterior density of the PMBM filter grows rapidly as more data is observed. Thus, approximation methods, like pruning and merging [8], [9], need to be used to keep the number of MBs at a tractable level, see [7], [10] for examples on how this can be achieved in extended target tracking. The computational cost of the PMBM filter can be greatly reduced by approximating the MB mixture (MBM) in the posterior as a single MB, which leads to the so-called Poisson MB (PMB) filter. Hence, a better trade-off between computational complexity and estimation performance may be obtained. A performance evaluation of filters based on different MB conjugate priors for point target estimation, presented in [5], has shown that the point target PMB filter has very promising performance in terms of estimation error and computational time. It is therefore of interest to develop a merging technique that can efficiently approximate the extended target PMBM posterior density as a PMB, leading to the extended target PMB filter.

In this paper, we focus on developing MB approximation methods, based on the extended target PMBM conjugate prior [6], [7], to estimate the current set of target states. Because each target can generate multiple measurements per time step, the data association problem becomes far more challenging in multiple extended target tracking, than it is in multiple point target tracking. Due to this difference between, and the different measurement models being used in, point target tracking and extended target tracking, the merging techniques developed for point targets [8], [9] cannot be directly applied to the extended target PMBM filter [6], [7] and adapting the point target MB merging techniques to extended targets is a non-trivial problem. In this paper, we first discuss the problems associated with applying the ideas from [8], [9] directly in an extended target tracking context, and then we propose several different methods to address these problems, specifically different merging techniques. This yields several variants of the extended target PMB filter, one for each merging technique.

The main contributions can be summarized as follows:

- In Section VI, we present the track-oriented PMB (TO-PMB) filter, which is an adaptation of the track-oriented marginal MeMBer-Poisson (TOMB/P) filter [8] for point targets, and analyze the drawbacks of this track-oriented merging approach.
- 2) In Section VII, we present the PMB filter using variational approximation. More specifically, we first apply the variational MB (VMB) algorithm [9] to form tracks

Y. Xia, K. Granström and L. Svensson are with the Department of Electrical Engineering, Chalmers University of Technology, Göteborg, Sweden. E-mail: firstname.lastname@chalmers.se.M. Fatemi is with Zenuity, Göteborg, Sweden. E-mail: maryam.fatemi@zenuity.com

<sup>&</sup>lt;sup>1</sup>In the context of MTT, multi-target conjugate prior was defined in [3] as "if we start with the proposed conjugate initial prior," then all subsequent predicted and posterior distributions have the same form as the initial prior".

hypothesizing potential detected targets<sup>2</sup>, i.e., existing tracks, and then we use related ideas to propose a greedy method to form new tracks. Two variants of the VMB algorithm are studied, one follows the method proposed in [9], and the other is inspired by the Set Joint Probabilistic Data Association (SJPDA) filter [11].

- In Section VIII, we present implementations of the proposed PMB filters for a common extended target model.
- 4) In Section IX, different variants of the extended target PMB filter are compared to the extended target PMBM filter, and to the extended target labelled MB (LMB) filter [12], in a thorough simulation study.

The remainder of the paper is organized as follows. In Section II, we introduce works related to this paper. In Section III, we introduce the background on Bayesian multi-target filtering, RFS modeling and the PMBM conjugate prior. In Section IV, we present the problem formulation. In Section V, we present the extended target PMB filter and identify the challenges that have to be solved for MBM merging. Nomenclature is presented in Table I.

# II. RELATED WORK

In this paper we derive a new tracking algorithm using RFS, hence we focus on RFS algorithms in the discussion of related work. Non-RFS extended target MTT algorithms include the JPDA-type approaches [13]–[15] and the Probabilistic Multiple Hypothesis Tracking approaches [16]–[18], see [19] for an elaborate overview of extended target tracking literature.

In some early works, RFS-based MTT approaches were developed based on principled approximations of posterior multitarget densities for both point and extended targets, including the Probability Hypothesis Density (PHD) filter [20]–[24], the Cardinalized PHD (CPHD) filter [25], [26] and the multi-target multi-Bernoulli (MeMBer) filter [27]. A key reason for the popularity of approaches based on principled approximations is that they sidestep the requirement to generate explicit associations of measurements to targets, though in extended target tracking different partitions of the set of measurements need to be computed prior to the approximations [21], [22], [26].

A multi-target conjugate prior gives an exact closed-form solution of the RFS-based multi-target Bayes filter. One main advantage of using conjugate priors in MTT is that the multi-target posterior density can be approximated arbitrarily well as long as sufficient parameters are used. Two types of multi-target conjugate priors can be found in the literature: the  $\delta$ -Generalized Labelled Multi-Bernoulli ( $\delta$ -GLMB) density [3], [12] and the PMBM density [7], [8], [28], [29]. A special case of a PMBM is a MBM density, which is obtained by setting the intensity of the Poisson Point Process (PPP) to zero in a PMBM [4], [30]. An MBM, in which each Bernoulli is uniquely labeled and have existence probability either 0 or 1, is similar in structure to the  $\delta$ -GLMB [4, Sec. IV].

#### TABLE I Nomenclature

- x: single target states; X: multi-target states; z: single measurement; Z, C: set of measurements.
- Blackboard bold letters, e.g., I, J, are used to represent set of indices. More specifically, J denotes the MB index set, I denotes the Bernoulli index set for the predicted MB, I<sup>j</sup> denotes the Bernoulli index set for the jth updated MB and Î denotes the Bernoulli index set for the approximating MB.
- Calligraphic letters, e.g., A, F, are used to represent spaces.
- | · |: set cardinality.
- det(X): determinant of matrix X.
- Tr(X): trace of matrix X.
- δ<sub>Y</sub>(X): a generalized Kronecker delta function that takes arbitrary arguments such as sets, vectors, etc., by

$$\delta_Y(X) \triangleq \begin{cases} 1, & \text{if } X = Y \\ 0, & \text{otherwise} \end{cases}$$
.

• 1<sub>Y</sub>(X): a generalized indicator function, by

$$1_{\mathbf{Y}}(\mathbf{X}) \triangleq \begin{cases} 1, & \text{if } \mathbf{X} \subseteq \mathbf{Y} \\ 0, & \text{otherwise} \end{cases}$$
.

- One-to-one mapping function  $\theta: X \to Y$ , such that  $\forall x, x' \in X, x \neq x' \Rightarrow \theta(x) \neq \theta(x')$ .
- $\langle a,b \rangle$ : inner product of a(x) and b(x),  $\langle a,b \rangle \triangleq \int a(x)b(x)dx$ .
- $\Pi_N$ : set of permutation functions on  $I_N \triangleq \{1, ..., N\}$

$$\Pi_N \triangleq \{\pi: I_N \to I_N | i \neq j \Rightarrow \pi(i) \neq \pi(j)\}.$$

- $\uplus$ : disjoint set union, i.e.,  $Y \uplus U = X$  means that  $Y \cup U = X$  and  $Y \cap U = \emptyset$ .
- $D_{\text{KL}}(p||q) \triangleq \int p(x) \log \left(\frac{p(x)}{q(x)}\right) dx$ : KL divergence between probability distributions p and q.
- $\Gamma_d(\cdot)$ : multivariate Gamma function.
- $\varphi_0(\cdot)$ : digamma function.
- $I_m$ : identity matrix of size  $m \times m$ .
- $\mathcal{GAM}(\gamma; a, b)$ : probability density function of Gamma distribution defined on  $\gamma > 0$  with shape a and rate b.
- \(\mathcal{TW}\_d(\chi, v, V)\): probability density function of inverse-Wishart
  distribution defined on positive definite \(d \times d\) matrix with degrees
  of freedom \(v\) and \(d \times d\) scale matrix \(V\).

The MBM is conjugate for MB and MBM birth; the  $\delta$ -GLMB is conjugate for labelled MB (mixture) birth; and the PMBM is conjugate for PPP birth. In the PMBM, this leads to two disjoint and independent sets of targets: targets that have been previously detected, and targets that exist but have never been detected [8]. The ability to form target trajectories using the PMBM conjugate prior can be achieved by performing MTT using sets of trajectories [28], [29], [31]. Using the  $\delta$ -GLMB conjugate prior, target trajectories are formed by connecting estimates from different times that have the same label whose uniqueness is ensured through the model [3], [12]. The relations between the PMBM and the  $\delta$ -GLMB multitarget conjugate priors were discussed in [4], [7], where it was shown that the PMBM density has a more efficient hypothesis structure than the  $\delta$ -GLMB density.

By explicitly capturing the associations of measurements, the PMBM filter [4], [6], [7] and the  $\delta$ -GLMB filter [12], [32]–[34] have been shown to outperform other RFS filters based on principled approximations for both point and extended targets, albeit with a higher computational cost. To reduce the computational complexity of the PMBM filter and the  $\delta$ -GLMB filter, a number of approximation methods have been developed. For instance, the LMB filter [12], [35] is

 $<sup>^2 \</sup>text{The term}$  potential target is used because Bernoulli densities may incorporate uncertain existence probability r, i.e., 0 < r < 1. Note that, this is different from the Bernoulli density in the GLMB representation, in which target existence probability is deterministic, i.e.,  $r \in \{0,1\}.$ 

an efficient approximation of the  $\delta$ -GLMB filter. Different variants of the PMB filter have been developed for point target filtering [8], [9], among which the VMB filter [9] has been shown to provide the most accurate target state estimation. The VMB filter operates by finding the best-fitting MB distribution that minimizes the KullbackLeibler (KL) divergence from the MBM in the true posterior. Simulation results have shown that RFS filters based on MB approximations [5], [8], [9], [12], [35] generally inherit the advantages of filters based on multitarget conjugate priors and outperform RFS filters based on principled approximations.

An MBM reduction technique is presented in [7] that merges MBs that 1) have the same number of Bernoullis, and 2) whose Bernoullis are pairwise similar in the sense of the KL-divergence. As a comparison, in this work we aim at merging the whole PMBM density as a single PMB density, which is a much more complicated problem, e.g., each MB in the MBM posterior may contain different number of Bernoullis. Also note that the filter implementation in [6], [7] is Hypothesis-Oriented, whereas in this paper we consider a Track-Oriented filter implementation.

#### III. BACKGROUND

In this section, we first give introductions to Bayesian filtering and RFS modeling. Next, we outline the multi-target transition model and the extended target measurement used in this work. Then, we present the PMBM conjugate prior for multiple extended target filtering.

# A. Bayesian multi-target filtering

In RFS-based MTT methods, target states and measurements are represented in the form of finite sets. Let  $\mathbf{x}_k$  denote the single target state at discrete time step k, and let  $\mathbf{X}_k$  denote the target set. In extended target tracking, the target state models both the kinematic properties, and the extent, of the target. The target set cardinality  $|\mathbf{X}_k|$  is a time-varying discrete random variable, and each target state  $\mathbf{x}_k \in \mathbf{X}_k$  is also a random variable. Further, let  $\mathbf{z}_k$  denote the single measurement at time step k, let  $\mathbf{Z}_k$  denote the set of measurements obtained at time step k, including clutter and target measurements, and let  $\mathbf{Z}^k$  denote the sequence of all the measurement sets received so far up to and till time step k.

The objective of multi-target filtering is to recursively compute the multi-target posterior density. Let  $f_{k|k}(\mathbf{X}_k|\mathbf{Z}^k)$ ,  $f_{k,k-1}(\mathbf{X}_k|\mathbf{X}_{k-1})$ , and  $f_k(\mathbf{Z}_k|\mathbf{X}_k)$  denote the multi-target set density, the multi-target transition density, and the multi-target measurement likelihood, respectively. The multi-target Bayes filter propagates in time the multi-target set density  $f_{k-1|k-1}(\mathbf{X}_{k-1}|\mathbf{Z}^{k-1})$  using the Chapman-Kolmogorov prediction

$$f_{k|k-1}(\mathbf{X}_k|\mathbf{Z}^{k-1}) = \int f_{k,k-1}(\mathbf{X}_k|\mathbf{X}_{k-1})f_{k-1|k-1}(\mathbf{X}_{k-1}|\mathbf{Z}^{k-1})\delta\mathbf{X}_{k-1}, \quad (1)$$

and the Bayes update

$$f_{k|k}(\mathbf{X}_k|\mathbf{Z}^k) = \frac{f_k(\mathbf{Z}_k|\mathbf{X}_k)f_{k|k-1}(\mathbf{X}_k|\mathbf{Z}^{k-1})}{\int f_k(\mathbf{Z}_k|\mathbf{X}_k)f_{k|k-1}(\mathbf{X}_k|\mathbf{Z}^{k-1})\delta\mathbf{X}_k}, \quad (2)$$

where the definition of set integral,  $\int f(\mathbf{X})\delta\mathbf{X}$ , can be found in [2, Section 3.3].

# B. Random set modeling

Two basic forms of RFS are the PPP and the Bernoulli process. The PPP density is given by

$$f(\mathbf{X}) = e^{-\mu} \prod_{\mathbf{x} \in \mathbf{X}} \mu f(\mathbf{x}), \tag{3}$$

and the PPP intensity  $D(\mathbf{x}) = \mu f(\mathbf{x})$  is determined by the scalar Poisson rate  $\mu$  and the spatial distribution  $f(\mathbf{x})$ . The Bernoulli density is given by

$$f(\mathbf{X}) = \begin{cases} 1 - r, & \mathbf{X} = \emptyset \\ rf(\mathbf{x}), & \mathbf{X} = \{\mathbf{x}\} \\ 0, & \text{otherwise} \end{cases}$$
 (4)

where  $r \in [0,1]$  is the probability of existence and  $f(\mathbf{x})$  is the existence-conditioned probability density function. The Bernoulli process offers a convenient way to capture the uncertainty regarding both target existence and target state.

Multiple independent targets can be represented as a multi-Bernoulli RFS X, which is a disjoint union of independent Bernoulli RFSs  $X^i$ , i.e.,  $X = \biguplus_{i \in \mathbb{I}} X^i$ , where  $\mathbb{I}$  denotes the Bernoulli index set. The RFS density of an MB process can be represented as

$$f(\mathbf{X}) = \begin{cases} \sum_{\mathbf{u}_{i \in \mathbb{I}} \mathbf{X}^{i} = \mathbf{X}} \prod_{i \in \mathbb{I}} f^{i}(\mathbf{X}^{i}), & |\mathbf{X}| \leq |\mathbb{I}| \\ 0, & |\mathbf{X}| > |\mathbb{I}| \end{cases}$$
 (5)

The MB density can be defined entirely by the parameters  $\{r^i, f^i(\cdot)\}_{i\in\mathbb{I}}$  of the involved Bernoulli RFSs. In MTT, different MBs typically correspond to the different data association sequences. A normalized, weighted sum of MB densities is called MBM, which can be defined entirely by the parameters

$$\{(W^j, \{r^{j,i}, f^{j,i}(\cdot)\}_{i \in \mathbb{I}^j})\}_{j \in \mathbb{J}},$$
 (6)

where  $\mathbb{J}$  is the MB index set;  $\mathbb{I}^j$  is the Bernoulli index set for the jth MB;  $r^{j,i}$  and  $f^{j,i}(\cdot)$  are the existence probability and existence-conditioned PDF of the ith Bernoulli in the jth MB, respectively;  $W^j$  is the normalized weight of the jth MB.

# C. Target transition and measurement models

- 1) Multi-target transition model: New targets appear independently of the targets that already exist. The target birth is assumed to be a PPP with intensity  $D^b(\mathbf{x})$ . A single target survives from time step k to time step k+1 with a probability of survival  $p^S(\mathbf{x}_k)$  at time step k. Targets evolve independently according to an i.i.d. Markov process with transition density  $f_{k+1,k}(\mathbf{x}_{k+1}|\mathbf{x}_k)$ .
- 2) Extended target measurement model: In extended target tracking, a common model to describe the number and the spatial distribution of generated measurements for each target is the inhomogeneous PPP [36]. The set of measurements  $\mathbf{Z}_k$  is a union of a set of clutter measurements and sets of target-generated measurements. The clutter is modeled as a PPP with Poisson rate  $\lambda$  and spatial distribution  $c(\mathbf{z})$ , and the clutter PPP intensity is  $\kappa(\mathbf{z}) = \lambda c(\mathbf{z})$ . The clutter measurements

are independent of targets and any target measurements. Each extended target may give rise to multiple measurements with a state dependent detection probability  $p^D(\mathbf{x}_k)$ . If the extended target is detected, the target measurements are modeled as a PPP with Poisson rate  $\gamma(\mathbf{x}_k)$  and spatial distribution  $\phi(\cdot|\mathbf{x}_k)$ , independent of any other targets and their corresponding generated measurements.

The measurement likelihood for a single extended target and a nonempty set of measurements **Z** is the product of the target detection probability and the PPP density of target-generated measurements

$$\ell_{\mathbf{Z}}(\mathbf{x}_k) = p^D(\mathbf{x}_k)e^{-\gamma(\mathbf{x}_k)} \prod_{\mathbf{z} \in \mathbf{Z}} \gamma(\mathbf{x}_k)\phi(\mathbf{z}|\mathbf{x}_k). \tag{7}$$

For an extended target state  $x_k$ , the effective detection probability is the product of target detection probability and the probability that target generates at least one measurement  $1 - e^{-\gamma(\mathbf{x}_k)}$ . Therefore the effective probability of missed detection is calculated as

$$q^{D}(\mathbf{x}_k) = 1 - p^{D}(\mathbf{x}_k) + p^{D}(\mathbf{x}_k)e^{-\gamma(\mathbf{x}_k)}.$$
 (8)

Note that the measurement likelihood for an empty measurement set,  $\ell_{\emptyset}(\mathbf{x}_k)$ , is also described by (8).

# D. PMBM conjugate prior

The PMBM conjugate prior was developed in [4], [8] for multiple point target filtering, and in [6], [7] for multiple extended target filtering. In the PMBM model, the target set is a union of two disjoint sets: the undetected targets  $X^u$  and the detected targets  $\mathbf{X}^d$ , i.e.,  $\mathbf{X} = \mathbf{X}^u \uplus \mathbf{X}^d$ . The distribution of targets that are undetected  $X^u$  is described by a PPP, while the distribution of targets that have been detected at least once  $\mathbf{X}^d$  is described by an MBM, independent of  $\mathbf{X}^u$ . The PMBM set density can be expressed as

$$f(\mathbf{X}) = \sum_{\mathbf{X}^u \uplus \mathbf{X}^d = \mathbf{X}} f^u(\mathbf{X}^u) \sum_{j \in \mathbb{J}} W^j f^j(\mathbf{X}^d), \quad (9a)$$

$$f^{u}(\mathbf{X}^{u}) = e^{-\langle D^{u}; 1 \rangle} \prod_{\mathbf{x} \in \mathbf{X}^{u}} D^{u}(\mathbf{x}), \tag{9b}$$

$$f^{u}(\mathbf{X}^{u}) = e^{-\langle D^{u}; 1 \rangle} \prod_{\mathbf{x} \in \mathbf{X}^{u}} D^{u}(\mathbf{x}), \tag{9b}$$

$$f^{j}(\mathbf{X}^{d}) = \sum_{\mathbf{y}_{i \in \mathbb{I}^{j}} \mathbf{X}^{i} = \mathbf{X}^{d}} \prod_{i \in \mathbb{I}^{j}} f^{j,i}(\mathbf{X}^{i}), \tag{9c}$$

and it can be defined entirely by the parameters,

$$D^{u}, \{(W^{j}, \{r^{j,i}, f^{j,i}(\cdot)\}_{i \in \mathbb{I}^{j}})\}_{j \in \mathbb{J}}.$$
 (10)

Because the PMBM density is a MTT conjugate prior, performing prediction (1) and update (2) means that we compute the predicted and updated, respectively, PMBM density parameters.

In the PMBM filter, each MB corresponds to a unique global hypothesis for the detected targets, i.e., a particular history of data associations for all measurements. Global hypotheses consist of single target hypotheses, each of which can incorporate a distribution of one of the different data association events, via a Bernoulli process. A track is defined as a collection of single target hypotheses corresponding to the same potential target<sup>3</sup> [8]. Given a predicted PMBM density in the update (2), for each predicted global hypothesis, there can be multiple possible data associations, each of which will result in an MB in the updated MBM.

# IV. PROBLEM FORMULATION

In the PMBM filter, without approximation, the number of MBs grows hyper-exponentially over time [4], [7]. In this work, we aim at developing an extended target filter that propagates a PMB density over time, i.e., a special case of the PMBM density (9) that has an MBM with a single MB. By only having a single MB describing detected targets, the number of parameters needed to be calculated in the prediction and update steps is reduced; as a result, the computational cost of the filter is greatly decreased.

If the posterior  $f_{k-1|k-1}(\cdot|\cdot)$  is a PMB density, then the predicted density  $f_{k|k-1}(\cdot|\cdot)$ , resulting from (1), is also PMB [8]. However, with a PMB prior  $f_{k|k-1}(\cdot|\cdot)$ , the Bayes update (2) results in a PMBM posterior  $f_{k|k}(\cdot|\cdot)$  due to the unknown data associations. The problem considered in this paper is how to approximate the true PMBM posterior  $f_{k|k}(\cdot|\cdot)$  with a PMB density  $f_{k|k}(\cdot|\cdot)$  to obtain a recursive PMB filter. The approximating PMB density consists of a disjoint union of a PPP  $\hat{f}_{k|k}^u(\mathbf{X}_k^u)$  and an MB  $g_{k|k}(\mathbf{X}_k^d)$  in the form of

$$\hat{f}_{k|k}(\mathbf{X}_k) = \sum_{\mathbf{X}_k^u \uplus \mathbf{X}_k^d = \mathbf{X}_k} \hat{f}_{k|k}^u(\mathbf{X}_k^u) g_{k|k}(\mathbf{X}_k^d). \tag{11}$$

In the approximation, our objective is to obtain a PMB that can retain as much information from the PMBM as possible. A natural choice for solving this problem is to minimize the KL divergence between the PMBM and the approximating PMB,

$$\min_{\hat{f}} D_{\mathrm{KL}}(f_{k|k}(\mathbf{X}_k)||\hat{f}_{k|k}(\mathbf{X}_k)), \tag{12}$$

where  $\hat{f}_{k|k}(\mathbf{X}_k)$  is restricted to a PMB. Because analytically minimizing the KL divergence (12) is intractable, we have to instead use approximations to obtain an efficient algorithm.

Note that the PPP describing the set of undetected targets does not have to be approximated [8], and that Bernoullis with small probability of existence can be efficiently approximated as being Poisson [38]. Thus, it is sufficient to consider approximating the MBM describing the set of detected targets as a PMB. The problem of approximating the MBM as a PMB is solved in two steps. We first consider methods for approximating the MBM as a single MB, and then we utilize the recycling method [38] to further approximate the approximating MB as a PMB. Nonetheless, analytically finding the MB density  $g_{k|k}(\mathbf{X}_k^d)$  that minimizes the KL divergence

$$\underset{g}{\operatorname{arg\,min}} \ D_{\mathrm{KL}}\bigg(\sum_{j\in\mathbb{J}} W^{j} f_{k|k}^{j}(\mathbf{X}_{k}^{d})||g_{k|k}(\mathbf{X}_{k}^{d})\bigg), \tag{13}$$

where  $\sum_{j\in\mathbb{J}}W^jf_{k|k}^j(\mathbf{X}_k^d)$  is an MBM, is still difficult.

Different approaches to efficiently approximate the MBM as a single MB has been well studied for point target filtering, see

<sup>&</sup>lt;sup>3</sup>The track defined here is different from the convention used in Multiple Hypothesis Tracker algorithms [37], where track is referred to as single trajectory hypothesis.

[8], [9]. In this work, we focus on developing MBM merging for extended targets in order to design an extended target PMB filter. One such approach is to use a method similar to the TOMB/P filter [8]. A more appealing approach is to use the variational method of [9] to search for the MB parameters that gives the smallest possible KL divergence (13). In this paper, we explore both alternatives.

# V. EXTENDED TARGET PMB FILTER

In this section, we first introduce the data association model, and then we present the extended target PMB filtering recursion, which includes prediction, update, MB approximation and recycling. In what follows, we omit explicit references to the time index k for brevity.

#### A. Data association

Let the predicted multi-target density be a PMB density with parameters

$$D^{u}, \{r^{i}, f^{i}(\cdot)\}_{i \in \mathbb{I}}, \tag{14}$$

let M be the index set of current measurement set Z, i.e.,  $\mathbf{Z} = \{\mathbf{z}^m\}_{m \in \mathbb{M}}$ , and let  $\mathcal{A}$  be the space of all data associations A given the predicted hypothesis. Given a set of predicted tracks, each of which corresponds to a single local hypothesis, and a set of measurements, we obtain one data association hypothesis for every combination of

- 1) a partition of the set of measurements into non-empty subsets called cells,
- 2) an assignment of each cell to either a predicted track, or to a potential new track.

In 2), each cell can be associated to only one track (predicted or new), and each predicted track can be associated either to one cell or to no cell. Cells that are not associated to a predicted track give rise to a new potential track.

In the paper, we express pairs of 1) and 2) as a partition of the set  $\mathbb{M} \uplus \mathbb{I}$ . Given the PMB density with parameters (14), an extended target data association  $A \in \mathcal{A}$  consists of a partition of  $\mathbb{M} \oplus \mathbb{I}$  into non-empty disjoint subsets called index cells  $\mathbb{C} \in A$ . An index cell can contain at most one Bernoulli index, and if the index cell C contains a Bernoulli index, let  $i_{\mathbb{C}}$  denote the corresponding Bernoulli index. Further, let  $\mathbf{C}_{\mathbb{C}}$ denote the measurement cell that corresponds to the index cell  $\mathbb{C}$ , i.e., the set of measurements  $\mathbf{C}_{\mathbb{C}} = \bigcup_{m \in \mathbb{C} \cap \mathbb{M}} \mathbf{z}^m$ . For any data association  $A \in \mathcal{A}$ , the likelihood of association  $L_A$  can be expressed as [10]

$$L_{A} = \prod_{\substack{\mathbb{C} \in A: \\ \mathbb{C} \cap \mathbb{I} = \emptyset \\ \mathbb{C} \cap \mathbb{M} \neq \emptyset}} L_{\mathbf{C}_{\mathbb{C}}}^{b} \prod_{\substack{\mathbb{C} \in A: \\ \mathbb{C} \cap \mathbb{I} \neq \emptyset \\ \mathbb{C} \cap \mathbb{M} \neq \emptyset}} L_{\mathbf{C}_{\mathbb{C}}}^{i_{\mathbb{C}}} \prod_{\substack{\mathbb{C} \in A: \\ \mathbb{C} \cap \mathbb{I} \neq \emptyset \\ \mathbb{C} \cap \mathbb{M} = \emptyset}}} L_{\mathbf{C}_{\mathbb{C}}}^{i_{\mathbb{C}}}, \quad (15a)$$

$$L_{\mathbf{C}_{\mathbb{C}}}^{b} = \begin{cases} \kappa^{\mathbf{C}_{\mathbb{C}}} + \langle D^{u}; \ell_{\mathbf{C}_{\mathbb{C}}} \rangle, & \text{if } |\mathbf{C}_{\mathbb{C}}| = 1 \\ \langle D^{u}; \ell_{\mathbf{C}_{\mathbb{C}}} \rangle, & \text{if } |\mathbf{C}_{\mathbb{C}}| \geq 1 \end{cases}. \quad (15b)$$

$$L_{\mathbf{C}_{\mathbb{C}}}^{b} = \begin{cases} \kappa^{\mathbf{C}_{\mathbb{C}}} + \langle D^{u}; \ell_{\mathbf{C}_{\mathbb{C}}} \rangle, & \text{if} \quad |\mathbf{C}_{\mathbb{C}}| = 1\\ \langle D^{u}; \ell_{\mathbf{C}_{\mathbb{C}}} \rangle, & \text{if} \quad |\mathbf{C}_{\mathbb{C}}| \ge 1 \end{cases} . \tag{15b}$$

$$L_{\mathbf{C}_{\mathbb{C}}}^{i} = r^{i} \langle f^{i}; \ell_{\mathbf{C}_{\mathbb{C}}} \rangle, \tag{15c}$$

$$L_{\emptyset}^{i} = 1 - r^{i} + r^{i} \langle f^{i}; q_{D} \rangle. \tag{15d}$$

$$L_{\emptyset}^{i} = 1 - r^{i} + r^{i} \langle f^{i}; q_{D} \rangle. \tag{15d}$$

The three products in (15a) correspond to:

• measurement cells that are associated to the background, i.e., either clutter or one of the previously undetected targets in the set  $X^u$ ,

- measurement cells that are associated to one of the previously detected targets in the set  $X^d$ ,
- previously detected targets  $X^d$  that are misdetected.

The weight of the global hypothesis, that resulted from the predicted global hypothesis with association  $A \in \mathcal{A}$ , is [7]

$$W_A = \frac{L_A}{\sum_{A \in A} L_A}.$$
 (16)

**Example 1.** Suppose that we have two measurements and one predicted track with corresponding index sets  $\mathbb{M} = (m_1, m_2)$ and  $\mathbb{I} = (i_1)$ , respectively. First, there are two possible ways to partition M into non-empty subsets:

- *single cell*:  $\{m_1, m_2\}$ ,
- two cells:  $\{m_1\}, \{m_2\}.$

In the first case, where we have a single cell, there are two possible ways to associate the cell and the predicted track:

- associate  $\{m_1, m_2\}$  to  $i_1$ ,
- associate  $\{m_1, m_2\}$  to the background, and  $i_1$  to a

In the second case, where we have two cells, there are three possible ways to associate the cells  $\{m_1\}$  and  $\{m_2\}$  and the

- associate  $\{m_1\}$  to  $i_1$  and  $\{m_2\}$  to the background,
- associate  $\{m_2\}$  to  $i_1$  and  $\{m_1\}$  to the background,
- associate both  $\{m_1\}$  and  $\{m_2\}$  to the background, and  $i_1$  to a misdetection.

Together this gives us five valid partitions of  $\mathbb{M} \oplus \mathbb{I}$ , i.e., five different data associations in A:

- $\begin{array}{l} \bullet \ A_1 = \big\{ \{m_1, m_2, i_1\} \big\}, \\ \bullet \ A_2 = \big\{ \{i_1\}, \{m_1, m_2\} \big\}, \\ \bullet \ A_3 = \big\{ \{m_1, i_1\}, \{m_2\} \big\}, \end{array}$

- $A_4 = \{\{m_2, i_1\}, \{m_1\}\},$   $A_5 = \{\{i_1\}, \{m_1\}, \{m_2\}\}$

In the PMBM conjugate prior, see (9), J contains indices for MBs that correspond to different possible sequences of data associations, from the first time step to the current time step. Without approximations, the number of hypotheses in J grows hyper-exponentially with time. In comparison, in a PMB filter the predicted distribution always contains a single global hypothesis and the number of hypotheses after the update (before merging and pruning) is therefore given by the number of possible pairs of 1) and 2). Expressions for the number of possible data associations in the PMBM filter are given in [10].

A complexity analysis of the extended target data association problem given in [7] shows that it is generally intractable to enumerate all the possible associations; thus approximations are needed for computational tractability. The prevailing approach to solving the data association problem is to truncate associations with negligible probability, i.e., associations A for which weight  $W_A$  (16) is approximately zero. In a recent work [10], a sampling based method is proposed, which directly maximizes the multi-target likelihood function and solves the data association problem in a single step.

# B. PMB prediction and update

Given a posterior PMB density with parameters

$$D^{u}, (r^{i}, f^{i}(\cdot))_{i \in \mathbb{I}}, \tag{17}$$

and the target transition model introduced in III-C1, the predicted density is a PMB density with parameters [8]

$$D_{+}^{u}, (r_{+}^{i}, f_{+}^{i}(\cdot))_{i \in \mathbb{T}},$$
 (18)

where

$$D_+^u(\mathbf{x}) = D^b(\mathbf{x}) + \langle D^u, p^S f_{k+1,k} \rangle, \tag{19a}$$

$$r_{+}^{i} = \langle f^{i}, p^{S} \rangle r^{i}, \tag{19b}$$

$$f_{+}^{i}(\mathbf{x}) = \frac{\langle f^{i}, p^{S} f_{k+1,k} \rangle}{\langle f^{i}, p^{S} \rangle}.$$
 (19c)

Given a PMB prior with parameters (18), a set of measurements **Z**, and the extended target measurement model introduced in III-C2, the updated density is a PMBM [7], with parameters

$$D^{u}, \{(W^{j}, \{r^{j,i}, f^{j,i}(\cdot)\}_{i \in \mathbb{I}^{j}})\}_{j \in \mathbb{J}},$$
 (20)

where the updated PPP intensity is

$$D^{u}(\mathbf{x}) = q^{D}(\mathbf{x})D_{\perp}^{u}(\mathbf{x}),\tag{21}$$

and the updated MBs resulting from the data associations are indexed by  $j \in \mathbb{J}$ . In other words, for each  $j \in \mathbb{J}$  there is a unique association  $A_j \in \mathcal{A}$ . The Bernoulli index set in the predicted MB is a subset of the Bernoulli index set in each updated MB, i.e.,  $\mathbb{I} \subseteq \mathbb{I}^j \ \forall \ j \in \mathbb{J}$ . In the MBM, Bernoullis with index  $i \in \mathbb{I}$  correspond to predicted tracks, and Bernoullis with index  $i \in \mathbb{I}^j \setminus \mathbb{I}$  correspond to new tracks.

For tracks updating detected targets, a hypothesis can be included either as a missed detection, or as an update using a measurement cell  $\mathbb{C}_{\mathbb{C}}$ . Consider a Bernoulli  $i \in \mathbb{I}$ , an association  $j \in \mathbb{J}$  and an index cell  $\mathbb{C} \in A_j$  such that  $\mathbb{C} = \{i\}$ , i.e.,  $C_{\mathbb{C}} = \emptyset$ . Then, under association j Bernoulli i is misdetected and the updated Bernoulli parameters are

$$r^{j,i} = \frac{r_+^i \langle f_+^i, q^D \rangle}{1 - r_+^i + r_+^i \langle f_+^i, q^D \rangle},$$
 (22a)

$$f^{j,i}(\mathbf{x}) = \frac{q^D(\mathbf{x})f_+^i(\mathbf{x})}{\langle f_+^i, q^D \rangle}.$$
 (22b)

Consider a Bernoulli  $i \in \mathbb{I}$ , an association  $j \in \mathbb{J}$  and an index cell  $\mathbb{C} \in A_j$  such that  $i \in \mathbb{C}$  and  $C_{\mathbb{C}} \neq \emptyset$ . Then, under association j Bernoulli i is detected by the set of measurements  $C_{\mathbb{C}}$  and the updated Bernoulli parameters are

$$r^{j,i} = 1, (23a)$$

$$f^{j,i}(\mathbf{x}) = \frac{\ell_{\mathbf{C}_{\mathbb{C}}}(\mathbf{x})f_{+}^{i}(\mathbf{x})}{\langle f_{+}^{i}, \ell_{\mathbf{C}_{\mathbb{C}}} \rangle}.$$
 (23b)

Lastly, consider an association  $j \in \mathbb{J}$  and an index cell  $\mathbb{C} \in A_j$  such that  $\mathbb{C} \cap \mathbb{I} = \emptyset$  (in this case  $C_{\mathbb{C}} \neq \emptyset$  by definition of the associations). Then, under association j the measurements  $C_{\mathbb{C}}$ 

are associated to the background (new target or clutter) and the parameters of the corresponding new Bernoulli are

$$r^{j,i} = \begin{cases} 1, & \text{if } |\mathbf{C}_{\mathbb{C}}| > 1\\ \frac{\langle D_{+}^{u}, \ell_{\mathbf{C}_{\mathbb{C}}} \rangle}{\kappa(\mathbf{C}_{\mathbb{C}}) + \langle D_{+}^{u}, \ell_{\mathbf{C}_{\mathbb{C}}} \rangle}, & \text{if } |\mathbf{C}_{\mathbb{C}}| = 1 \end{cases}$$
(24a)

$$f^{j,i}(\mathbf{x}) = \frac{\ell_{\mathbf{C}_{\mathbb{C}}}(\mathbf{x})D_{+}^{u}(\mathbf{x})}{\langle D_{+}^{u}, \ell_{\mathbf{C}_{\mathbb{C}}} \rangle}.$$
 (24b)

After updating, the number of Bernoullis representing previously detected targets in each MB equals the number of Bernoullis in the predicted MB, while the number of Bernoullis representing new potentially detected targets becomes  $|\mathbb{I}^j \setminus \mathbb{I}|$  in the jth MB.

The PMB prediction and update steps are in closed forms and their derivations can be found in [7], [8]. Also, note that the PMB update equations (22), (23) and (24) also hold for the standard measurement model, in which at most one measurement can be associated to a target, i.e.,  $|\mathbf{C}_{\mathbb{C}}| \leq 1$ , used in point target tracking.

C. MB approximation and its challenges

Each MB in the MBM,

$$W^{j}, \{r^{j,i}, f^{j,i}(\cdot)\}_{i \in \mathbb{I}^{j}}, j \in \mathbb{J}, \tag{25}$$

represents a global hypothesis of a collection of possible independent potential targets, and a relative likelihood of this hypothesis. Each potential target hypothesized to exist is described by a Bernoulli, with parameters

$$\{r^{j,i}, f^{j,i}(\cdot)\}, i \in \mathbb{I}^j, j \in \mathbb{J}. \tag{26}$$

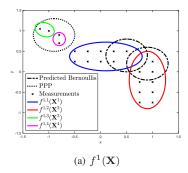
The objective of the MB approximation is to represent a collection of possible independent targets with a single MB, by exploiting the information contained in all the uncertain hypotheses. More specifically, we would like to obtain the MB  $g(\mathbf{X})$  with parameters

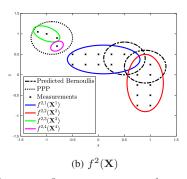
$$\{\hat{r}^{\iota}, g^{\iota}(\cdot)\}_{\iota \in \hat{\mathbb{I}}},\tag{27}$$

where  $\hat{\mathbb{I}}$  denotes the index set of the approximating Bernoullis, that best matches the MBM  $f(\mathbf{X})$  with parameters (26) using some merging techniques.

We identify that there are three major challenges related to the MB approximation. The first challenge is to determine the number of Bernoullis  $|\hat{\mathbb{I}}|$  in the approximating MB, since different MBs  $f^j(\mathbf{X})$  may not all contain the same number of Bernoullis  $|\mathbb{I}^j|$ . Given  $|\mathbb{M}|$  measurements, there are  $2^{|\mathbb{M}|}-1$  possible ways in which we can form a subset of measurements. Each such subset can, in an association, be associated to a new potentially detected target. As a result, different global hypotheses may contain a different number of new tracks. Therefore determining how many new tracks should be formed in the MB approximation is a challenge.

In this work, we choose to approximate the MBM as a single MB by merging. Then, the second challenge is to determine, for each  $\iota \in \hat{\mathbb{I}}$ , which Bernoullis  $f^{j,i}(\mathbf{X})$  should be merged to form  $\hat{r}^{\iota}$  and  $g^{\iota}(\mathbf{x})$ . Each MB is invariant to any ordering





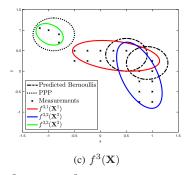


Fig. 1. Illustrative example with three MBs  $f^1(\mathbf{X})$ ,  $f^2(\mathbf{X})$  and  $f^3(\mathbf{X})$ , with weights  $W^1 = 0.4$ ,  $W^2 = 0.35$ ,  $W^3 = 0.25$ . Densities are presented by their level curves. Bernoullis marked in the same color red/blue correspond to the measurement update of the same track. Bernoullis marked in red and blue stand for the updates of previously detected targets, whereas Bernoullis marked in green and magenta stand for the first detections of undetected targets.

of its Bernoullis, so determining how to select Bernoullis to merge across different MBs is a challenge<sup>4</sup>.

The third challenge is to determine how to merge the selected Bernoullis into a single Bernoulli. Compared to the first two challenges, the third challenge is less difficult, e.g., given a Bernoulli mixture, one can take the weighted sum of different parameters. In extended target filtering, the merging of different target state densities is typically addressed by analytically minimizing the KL divergence, see, e.g., [39], [40]. In Section VII, we discuss some implementation details regarding Bernoulli merging under a common extended target model, called Gamma Gaussian inverse Wishart (GGIW) [26].

We illustrate these three major challenges with the following example.

**Example 2.** Consider a two-dimensional scenario shown in Fig. 1, with an MBM posterior that contains three global hypotheses. For simplicity, we assume that each Bernoulli has existence probability equal to one. In this scenario, two targets detected for the first time are closely spaced, and therefore it is ambiguous whether there are two closely spaced new targets or one bigger new target.

The first challenge is determining how to choose the number of Bernoullis in the approximating MB. In this example, it may be reasonable to have two Bernoullis corresponding to the two predicted tracks. However, since the new tracks in the three global hypotheses are not all the same, it is nontrivial to select the total number of Bernoullis to include in the approximating density.

The second challenge is determining how to select the Bernoulli components to be merged. For example, when creating approximating Bernoullis for existing tracks, it could then make sense to merge  $f^{1,1}(\mathbf{X}^1)$ ,  $f^{2,1}(\mathbf{X}^1)$  and  $f^{3,2}(\mathbf{X}^2)$  into one Bernoulli and  $f^{1,2}(\mathbf{X}^2)$ ,  $f^{2,2}(\mathbf{X}^2)$  and  $f^{3,1}(\mathbf{X}^1)$  into a second, but it is generally non-trivial to select which Bernoullis, in the different hypotheses, to merge. The problem becomes more complicated when we consider the approximation of new tracks, since the number of approximating

Bernoullis representing new potentially detected targets is yet to be decided.

The third challenge is determining how to merge the selected Bernoullis. For example, if we choose to merge  $f^{1,1}(\mathbf{X}^1)$ ,  $f^{2,1}(\mathbf{X}^1)$  and  $f^{3,2}(\mathbf{X}^1)$ , then we should seek for accurate merging techniques so that the approximating Bernoulli can retain as much information as possible from these three Bernoullis.

# D. Recycling

For the approximating MB  $g(\mathbf{X})$  with parameters (27), the recycling method of [38] can be applied to Bernoullis with low existence probability  $\hat{r}^{\iota}$ . The recycled components are approximated as being Poisson and are incorporated into the PPP representing undetected targets for generating possible new targets in subsequent steps. After recycling, the total PPP intensity of the set of undetected targets can be expressed as

$$\hat{D}^{u}(\mathbf{x}) = D^{u}(\mathbf{x}) + \sum_{\iota \in \hat{\mathbb{I}}: \hat{r}^{\iota} < \tau_{r}} \hat{r}^{\iota} g^{\iota}(\mathbf{x}), \tag{28}$$

where  $\tau_r$  is a threshold, and a typical choice is  $\tau_r = 0.1$ . The benefits of recycling in the point target PMBM and PMB filters have been discussed in [5], [38]. In this work, we utilize MB approximation methods and recycling to approximate the PMBM posterior density as a PMB.

# VI. TRACK-ORIENTED PMB FILTER

In this section, we seek to adapt the merging technique used in the point target TOMB/P filter [8] to extended targets, and we call the resulting algorithm TO-PMB. Following a similar track-oriented merging approach as in [8], single target hypotheses updating different tracks are assumed to be independent, and hypotheses in the same track are merged into a single Bernoulli across different global hypotheses. We also discuss the extended target TO-PMB filter's relation to the extended target LMB filter [12], and the drawbacks of this track-oriented merging approach.

<sup>&</sup>lt;sup>4</sup>In this work, we choose not to merge Bernoullis within the same MB. An MB corresponds to a collection of independent potential targets. It is generally inappropriate to represent two or more independent targets with a single Bernoulli.

# A. Track formation

In the TOMB/P filter for point targets [8], in each update a new track is initiated for each measurement. In extended target tracking, the data association includes partitioning of the measurements, see Section V-A. Thus, the extended target equivalent to initiating a new track for each measurement, is to initiate a new track for each unique cell in all the measurement partitions.

We discuss the formations of new tracks and existing tracks separately. For predicted tracks ( $\iota \in \mathbb{I}$ ), the approximating Bernoulli  $g^{\iota}(\mathbf{X}^{\iota})$  is formed by merging all Bernoullis that correspond to the same track in different global hypotheses:

$$\underset{g^{\iota}}{\operatorname{arg\,min}} \ D_{\mathrm{KL}}\bigg(\sum_{j\in\mathbb{J}} W^{j} f^{j,\iota}(\mathbf{X}^{\iota})||g^{\iota}(\mathbf{X}^{\iota})\bigg), \tag{29}$$

where  $f^{j,i}(\mathbf{X}^i)$  has existence probability and existence-conditioned PDF in the form of (22) or (23). The explicit expression of the approximating single Bernoulli  $g^{\iota}(\mathbf{X}^{\iota})$  is given in Appendix A.

The number of new tracks is determined by the number of different measurement cells associated to the background in different global hypotheses. Mathematically, each index cell  $\mathbb{C} \in \bigcup_{j \in \mathbb{J}} A_j$  such that  $\mathbb{C} \cap \mathbb{I} = \emptyset$ , creates a new track. For a data association  $j \in \mathbb{J}$ , and a new Bernoulli  $i \in \mathbb{I}^j \setminus \mathbb{I}$ , we define the surjective mapping  $\phi:(j,i)\to \hat{\mathbb{I}}\setminus\mathbb{I}$ , which gives the index of the corresponding Bernoulli in the PMB approximation. Consider two MBs  $j_1, j_2 \in \mathbb{J}$ , and Bernoulli  $i_1$  from MB  $j_1$ , and Bernoulli  $i_2$  from MB  $j_2$ , where  $i_1,i_2\in$  $\mathbb{I} \setminus \mathbb{I}$ , i.e., both Bernoullis represent new tracks. Let  $\mathbb{C}_{j_1,i_1}$  and  $\mathbb{C}_{j_2,i_2}$  be the corresponding index cells. If  $\mathbb{C}_{j_1,i_1}=\mathbb{C}_{j_2,i_2}$ , then both Bernoullis were created by the same measurements, and we have that  $\phi((j_1, i_1)) = \phi((j_2, i_2))$ . For new tracks  $(\iota \in \hat{\mathbb{I}} \setminus \mathbb{I})$ , the approximating Bernoulli is formed by merging single target hypotheses updated by the same measurement cell across different global hypotheses, according to

$$\underset{g^{\iota}}{\operatorname{arg\,min}} \ D_{\mathrm{KL}} \bigg( \sum_{j \in \mathbb{J}} W^{j} \sum_{i \in \mathbb{J}^{j}} \delta_{\iota}(\phi((j,i))) f^{j,i}(\mathbf{X}^{i}) || g^{\iota}(\mathbf{X}^{\iota}) \bigg).$$
(30)

# B. Relation to the LMB filter

It was shown in [4] that the  $\delta$ -GLMB density is analogous to a type of MBM with deterministic Bernoulli existence probability on a labelled state space, and that labels are not required for the conjugacy property of the MBM. Similar to the merging strategy used in TO-PMB, in the LMB approximation Bernoullis with the same label, viz. updating the same track, are merged across different MBs [12].

The main difference between the proposed TO-PMB filter and the extended target LMB filter lies in the formation of new tracks, which is due to the different birth models used in these two filters; in the latter, the number of new tracks is determined by the number of Bernoulli components in the labelled MB birth density. A discussion on the choice of birth model, either Poisson or multi-Bernoulli, can be found in [30].

For the extended target LMB filter implementation proposed in [12], the conversion from LMB density to GLMB density

is needed before performing the update step, whereas in the extended target PMB filter, there is no need to perform such type of conversion since the PMB density is, indeed, a special case of PMBM with a single MB component. Note that the LMB-to-GLMB conversion in the LMB filter can be circumvented by considering the LMB mixture form with probabilistic Bernoulli existence probability in the MB update step, and a similar idea has been applied in [41] for point target tracking.

# C. Drawbacks of track-oriented merging

The track-oriented merging approach is simple to implement; nevertheless, it has several drawbacks. In the LMB filter and the TO-PMB filter, tracks are approximated as independent. This assumption holds well for the case that targets are well separated. However, the dependency between targets becomes inescapable when targets are closely spaced. A direct result is that both the LMB filter and the TO-PMB filter suffer from performance degradation when coalescence happens, see [5] and [8] respectively, for their performance evaluation on point target filtering.

In addition to the above drawback, the track-oriented merging approach is particularly unfit for the extended target PMB filter, in which new tracks created by different measurement cells are approximated as independent. If some of the measurement cells have shared measurements, which is the typical case, the new tracks can be highly dependent since they never co-exist in the same data association hypothesis, and approximating such tracks as independent may yield large errors. Also, creating as many new tracks as there are measurement cells often yields intractably many Bernoullis (tracks) with low existence probabilities.

Remark: It is also theoretically possible to develop a measurement cell-oriented extended target PMB filter by adapting the measurement-oriented merging technique used in the measurement-oriented marginal MeMBer-Poisson point target filter [8]. The idea is to create a track for each measurement cell collecting all single target hypotheses updated by the measurement cell, and a track for each prior track containing only the missed detection hypothesis. Similarly to track-oriented merging, the drawback of this measurement cell-oriented merging technique is that creating a track for each measurement cell may yield intractably many Bernoullis with low existence probabilities.

# VII. VARIATIONAL MULTI-BERNOULLI FILTER

As discussed, approximating the MBM using the trackoriented merging approach is likely to yield large estimation errors. The main issue of this problem lies in approximating (highly) dependent new tracks as independent. This gives rise to the need to develop a more appropriate method for approximating new tracks. In this section, we investigate merging techniques that are more accurate and yield fewer tracks, thereby simultaneously mitigating these weaknesses. More specifically, we divide the MB approximation into two separate parts: formation of new tracks and formation of existing tracks. The premise of this implementation is that the existing tracks and new tracks are assumed to be independent. Although there exist situations where this approximation is less accurate, it is essential for our approach to obtain a tractable solution. The following theorem shows that breaking down the KL minimization problem of (13) into two separate KL minimization subproblems yields intuitive outcomes.

**Theorem 1.** Suppose that we have a mixture density, where each mixture component consists of two independent multitarget densities,

$$f(\mathbf{X}) = \sum_{j \in \mathbb{J}} W_j \sum_{\mathbf{X}^1 \uplus \mathbf{X}^2 = \mathbf{X}} f^j(\mathbf{X}^1) f^j(\mathbf{X}^2), \quad (31)$$

Also, suppose we wish to find an approximate density,

$$g(\mathbf{X}) = \sum_{\hat{\mathbf{X}}^1 \uplus | \hat{\mathbf{X}}^2 = \mathbf{X}} g^1(\hat{\mathbf{X}}^1) g^2(\hat{\mathbf{X}}^2), \tag{32}$$

that minimizes the KL divergence

$$\min_{g^1, g^2} D_{KL}(f(\mathbf{X})||g(\mathbf{X})). \tag{33}$$

It holds that

$$\min_{g^1,g^2} D_{KL}(f(\mathbf{X})||g(\mathbf{X})) \leq \qquad \qquad \text{By simplifying the MB set integral in (35)} \\
\min_{g^1} D_{KL} \left( \sum_{j \in \mathbb{J}} W_j f^j(\mathbf{X}^1) \Big| \Big| g^1(\hat{\mathbf{X}}^1) \right) + \qquad \qquad \text{proximate upper bound of the objective (35) cataline as [9, Section III.A]} \\
\min_{g^2} D_{KL} \left( \sum_{j \in \mathbb{J}} W_j f^j(\mathbf{X}^2) \Big| \Big| g^2(\hat{\mathbf{X}}^2) \right). \quad (34) \qquad D_{UB}(f(\mathbf{X})||g(\mathbf{X})) = -\sum_{j \in \mathbb{J}, \pi \in \Pi_N^j} W^j q^j(\pi^j)$$

The proof of this theorem is given in Appendix B.

By applying Theorem 1, the KL divergence in (33) can be approximately minimized by finding its minimum upperbound which consists of two parts: one is the KL divergence between the marginal density of the existing tracks  $\sum_{i\in\mathbb{I}}W_jf^{j,i}(\mathbf{X}^i), i\in\mathbb{I}$  and its approximating MB density, and the second is the KL divergence between the marginal density of the new tracks  $\sum_{j\in\mathbb{J}} W_j f^{j,i}(\mathbf{X}^i), i\in\hat{\mathbb{I}}\setminus\mathbb{I}$  and its approximating MB density.

For approximating the existing tracks  $\iota \in \mathbb{I}$ , we employ the variational merging technique in [9], which can yield more accurate merging results than the track-oriented merging approach, especially in scenario with coalescence. Two variants of the VMB algorithm are studied. We first review the VMB algorithm implemented in [9], and then we study an alternative implementation of this algorithm that is inspired by the SJPDA filter [11]. For approximating the new tracks  $\iota \in \hat{\mathbb{I}} \setminus \mathbb{I}$ , we propose a greedy method to merge similar Bernoullis (tracks), also in the sense of minimizing the KL divergence.

#### A. Existing track formation

Because the number of existing tracks remains the same after updating, the VMB algorithm, proposed in [9] for point target filtering<sup>5</sup>, can be applied to the existing track approximation in the extended target PMB filter. In the VMB algorithm, the objective is to obtain an approximate MB  $q(\mathbf{X})$ that minimizes the KL divergence:

$$\min_{g} D_{KL}(f(\mathbf{X})||g(\mathbf{X})) = \min_{g} - \int f(\mathbf{X}) \log g(\mathbf{X}) \delta \mathbf{X}, (35)$$

where  $f(\mathbf{X})$  is an MBM describing the existing tracks, with parameters (6). An approximate solution of (35) is based on minimizing the upper bound of the true objective, following a similar process to Expectation-Maximization [42].

Note that due to the set representation of the multi-target posterior density, in the MB approximation it neither has to be the case that Bernoullis updating the same track are merged, nor does it have to be the case that Bernoullis updated by the same measurement cell are merged. Instead, we should treat the correspondences between the Bernoullis in  $f(\mathbf{X})$  and the Bernoullis in  $g(\mathbf{X})$  as missing data  $g(\pi)$ which probabilistically determines which Bernoullis in  $f(\mathbf{X})$ should be merged for each approximating Bernoulli in  $g(\mathbf{X})$ . Solving the optimization problem (35) can then be interpreted as finding the best missing data distribution  $\hat{q}(\pi)$ .

By simplifying the MB set integral in (35) into a series of Bernoulli integrals, and using the log-sum inequality, an approximate upper bound of the objective (35) can be expressed as [9, Section III.A]

$$D_{\text{UB}}(f(\mathbf{X})||g(\mathbf{X})) = -\sum_{j \in \mathbb{J}, \pi \in \Pi_N^j} W^j q^j(\pi^j)$$

$$\times \sum_{\iota \in \mathbb{J}} \int f^{j,\pi^j(\theta(\iota))}(\mathbf{X}) \log g^{\iota}(\mathbf{X}) \delta \mathbf{X}, \quad (36)$$

where  $N = |\mathbb{I}|$  denotes the number of Bernoullis that correspond to existing tracks in each MB of f(X) and the approximating MB  $g(\mathbf{X})$ ;  $\Pi_N^{\mathfrak{I}}$  is the set of all ways to assign the Bernoullis  $f^{j,i}(\mathbf{X}^i), i \in \mathbb{I}, j \in \mathbb{J}$  to the Bernoullis  $q^{\iota}(\mathbf{X}^{\iota}), \iota \in \mathbb{I}$ ; the missing data  $q^{j}(\pi^{j})$  is constrained to vary only with the jth MB, and satisfies  $q^{j}(\pi^{j}) \geq 0$  and  $\sum_{\pi^j \in \Pi^j} q^j(\pi^j) = 1$ . The standard method for optimizing (36) is block coordinate descent, which alternates between minimization with respect to  $g^{\iota}(\mathbf{X}^{\iota})$  (M-step) and  $q^{j}(\pi^{j})$ (E-step). Note that the block coordinate descent algorithm is guaranteed to converge to a certain value because the minimization objective (cross entropy) is always non-negative (and thus finite).

Because solving the minimization problem (36) suffers from combinatorial complexity, approximation is needed to obtain a tractable solution. Here, two different approximation methods are studied. The first one follows the method proposed in [9], and the second is inspired by the SJPDA filter [11]

1) Efficient approximation of feasible set: Because the minimization problem of (36) involves missing data  $q^{j}(\pi^{j})$ for every MB  $f^{j}(\mathbf{X})$ , a simplified missing data representation is desirable. The minimization of the upper bound (36) can be solved approximately as [9, Appendix A.C]

$$\underset{q(h,\iota)\in\mathcal{M}}{\operatorname{argmin}} - \sum_{\iota\in\mathbb{T}} \int \left(\sum_{h\in\mathbb{H}} q(h,\iota) f^{h}(\mathbf{X})\right) \log g^{\iota}(\mathbf{X}) \delta \mathbf{X}, \quad (37)$$

<sup>&</sup>lt;sup>5</sup>In the point target filtering, each measurement creates a new track, hence, different global hypotheses consist of the same number of tracks.

where  $\mathbb{H}$  is the index set of single target hypotheses included in the global hypotheses,  $q(h, \iota)$  specifies the weight of Bernoulli  $f^h(\mathbf{X}^h)^6$  in  $f(\mathbf{X})$  that is assigned to the approximating Bernoulli  $g^{\iota}(\mathbf{X}^{\iota})$ , and the feasible set  $\mathcal{M}$  is an approximation needed for tractability [9, Section III.C]

$$\mathcal{M} = \left\{ q(h, \iota) \ge 0 \middle| \sum_{h \in \mathcal{H}} q(h, \iota) = 1 \ \forall \ \iota \in \mathbb{I}, \right.$$
$$\left. \sum_{\iota \in \mathbb{I}} q(h, \iota) = p_h \ \forall \ h \in \mathbb{H} \right\}. \quad (38)$$

The constraint  $p_h$  satisfies  $p_h = \sum_{\iota \in \mathbb{I}} p_{\iota}(h)$ , where

$$p_{\iota}(h) = \sum_{j \in \mathbb{I}} W^{j} \delta_{f^{j,\theta(\iota)}(\mathbf{X}^{\theta(\iota)})}(f^{h}(\mathbf{X}^{h})), \iota \in \mathbb{I}$$
 (39)

Note that here the missing data distribution is no longer constrained to vary only with the global hypotheses. In this case, each approximating Bernoulli can be expressed as the weighted sum of different single target hypothesis densities, and the M-step becomes

$$\underset{g^{\iota}}{\operatorname{arg\,min}} \ D_{\mathrm{KL}}\bigg(\sum_{h\in\mathbb{H}} q(h,\iota)f^{h}(\mathbf{X}^{h})\bigg|\bigg|g^{\iota}(\mathbf{X}^{\iota})\bigg), \qquad (40)$$

while the E-step reverts to:

$$\underset{q(h,\iota)}{\operatorname{argmin}} \sum_{h \in \mathbb{H}} \sum_{\iota \in \mathbb{I}} -q(h,\iota) \int f^{h}(\mathbf{X}) \log g^{\iota}(\mathbf{X}) \delta \mathbf{X}, \qquad (41)$$

$$\text{subject to } \sum_{\iota \in \mathbb{I}} q(h,\iota) = p_{h} \ \forall \ h,$$

$$\sum_{h \in \mathbb{H}} q(h,\iota) = 1 \ \forall \ \iota,$$

$$q(h,\iota) > 0 \ \forall \ h,\iota.$$

Problems of this type can be solved using methods such as the simplex algorithm [43].

The VMB algorithm based on efficient approximation of feasible missing data set, abbreviated as EAFS-VMB, can be initialized with the marginal association probabilities (39). Although exact calculation of these quantities is intractable for a data association problem with combinatorial complexity, we can obtain approximate estimates by only considering truncated global hypotheses with non-negligible weights using approximation methods including gating/clustering, partitioning and assignment.

2) Most likely assignment: The number of MBs in the MBM can be kept at a tractable level after truncating the global hypotheses with negligible weights. This allows us to study an alternative approach to solving the minimization problem of (36), following a similar approach to the KL-SJPDA [11]. The KL-SJPDA filter with known number of targets seeks to find the ordered distribution in the same unordered family, such that the new density can be most accurately approximated

with a Gaussian density, in the KL sense. Since in the PMBM conjugate prior form each global hypothesis consists of the same number of existing tracks and the multi-target RFS density is order invariant, similar ideas in KL-SJPDA can be applied here to find the best-fitting MB.

Let us revisit the approximating upper bound (36) that we want to minimize. Suppose that Bernoullis in different MBs are indexed by the same superscript i if they correspond to single target hypotheses updating the same track, and that only Bernoullis with the same superscript i can be merged. Because the approximate MB density is invariant to the indexing of the Bernoullis it contains, the selection of the assignment mapping  $\pi$  in each MB will not change the MBM  $f(\mathbf{X})$ , but will only determine which Bernoullis are going to be merged. The minimization problem can be interpreted as the permutation of the Bernoullis in each MB in such a manner that the upper bound (36) is minimized, but the density of the reordered  $f(\mathbf{X})$  remains unchanged.

We choose to find the single most likely assignment  $\hat{\pi}^j$  for each MB  $f^j(\mathbf{X})$ . The resulting algorithm is abbreviated as MLA-VMB. In this case, the missing data distribution under each MB is a point mass, i.e.,  $q^j(\hat{\pi}^j) = 1$ , and the the minimization of (36) with respect to the missing data distribution can be expressed as

$$\hat{\pi}^{j} = \underset{\pi^{j}}{\operatorname{argmin}} - \sum_{\iota \in \mathbb{I}} \int f^{j,\pi^{j}(\theta(\iota))}(\mathbf{X}) \log g^{\iota}(\mathbf{X}) \delta \mathbf{X}, j \in \mathbb{J},$$
(42)

where the most likely assignment  $\hat{\pi}^j$  can be obtained using 2D assignment algorithm such as the Hungarian algorithm [44]. The minimization of (36) with respect to the approximating MB  $g(\mathbf{X})$  simplifies to

$$\underset{g^{\iota}}{\operatorname{arg\,min}} \ D_{\mathrm{KL}}\bigg(\sum_{j\in\mathbb{J}} W^{j} f^{j,\hat{\pi}^{j}(\theta(\iota))}(\mathbf{X}^{\hat{\pi}^{j}(\theta(\iota))}) \Big| \Big| g^{\iota}(\mathbf{X}^{\iota})\bigg). \tag{43}$$

This means that each approximating Bernoulli in  $g(\mathbf{X})$  can be obtained by merging Bernoullis in  $f(\mathbf{X})$  with the same assignment mapping.

3) Illustration: It can be noticed that, in the first iteration of the VMB algorithm, both the M-step (40) in the EAFS-PMB, and the M-step (43) in the MLA-VMB, are equivalent to the MBM merging step (29) used in the TO-PMB filter. This means that the VMB algorithm can be considered as an improvement on the track-oriented merging approach used in the TO-PMB filter.

We illustrate how the assignment mapping in each MB changes in each iteration of the MLA-VMB, and how the assignment weight matrix changes in each iteration of the EAFS-VMB with the following example.

**Example 3.** Consider the same scenario illustrated in Fig. 1. For the TO-PMB filter, Bernoullis updating the same target will be merged, i.e.,  $f^{1,1}(\mathbf{X}^1)$ ,  $f^{2,1}(\mathbf{X}^1)$  and  $f^{3,1}(\mathbf{X}^1)$  will be merged to obtain  $g^1(\mathbf{X}^1)$ ;  $f^{1,2}(\mathbf{X}^2)$ ,  $f^{2,2}(\mathbf{X}^2)$  and  $f^{3,2}(\mathbf{X}^2)$  will be merged to obtain  $g^2(\mathbf{X}^2)$ .

For the MLA-VMB, we can permute the order of Bernoullis and find the optimal permutation for each MB according to the E-step (42). Assume that, in this case, Bernoulli density

<sup>&</sup>lt;sup>6</sup>The same single target hypothesis may be included in different global hypotheses, thus it satisfies that  $|\mathbb{H}| \leq \sum_{j \in \mathbb{J}} |\mathbb{I}^j|$ . We use the single superscript h to denote the index of a single target hypothesis density in order to distinguish it from the double superscript (j,i), which denotes the indices of Bernoullis in the MBM.

# TABLE II PSEUDO CODE OF NEW TRACKS FORMING

```
Input: \{(W^j, \{f^{j,i}(\mathbf{X}^i)\}_{i\in\mathbb{I}^j\setminus\mathbb{I}})\}_{j\in\mathbb{J}}, merging threshold \tau_n
Output: \{g^{\iota}(\mathbf{X}^{\iota})\}_{\iota \in \hat{\mathbb{I}} \setminus \mathbb{I}}
    Sort f^{j}(\mathbf{X}) in the descending order of W^{j};
     I \leftarrow 0; \hat{\mathbb{I}} \leftarrow \mathbb{I};
    for all j = \{1, ..., |J|\} do
          if |\mathbb{I}^j| > |\mathbb{I}| then
                for all i = \{ |\mathbb{I}| + 1, ..., |\mathbb{I}^j| \} do
                      if (j,i) \notin \mathbb{L}^l \ \forall \ l = \{1,...,I\} then
                           I \leftarrow I + 1; \mathbb{L}^{I} \leftarrow \mathbb{L}^{I} \cup \{(j,i)\}; \hat{\mathbb{I}} \leftarrow \hat{\mathbb{I}} \cup \{|\mathbb{I}| + I\};
                      for all j^+ = \{j+1,...,|\mathbb{J}|\} do
                           if |\mathbb{I}^{j^+}| > |\mathbb{I}| then
                                 for all i^+ = \{ |\mathbb{I}| + 1, ..., |\mathbb{I}^{j^+}| \} do
                                       if (j,i) \notin \mathbb{L}^l \ \forall \ l \in \{1,...,I\} then
                                             d^{i^+} \leftarrow D_{\text{SKL}}(f^{j,i}(\mathbf{X})||f^{j^+,i^+}(\mathbf{X}));
                                        end if
                                  end for
                                  [d^*, i^*] \leftarrow \min(d);
                                 if d^* < \tau_n then \mathbb{L}^I \leftarrow \mathbb{L}^I \cup \{(j^+, i^*)\};
                                  end if
                            end if
                      end for
                end for
          end if
    end for
    for all \iota = \{ |\mathbb{I}| + 1, ..., |\hat{\mathbb{I}}| \} do
          Calculate g^{\iota}(\mathbf{X}^{\iota}) using (44).
```

 $f^{3,2}(\mathbf{X}^2)$  is closer to  $g^1(\mathbf{X}^1)$  than  $g^2(\mathbf{X}^2)$ , and that Bernoulli density  $f^{3,1}(\mathbf{X}^1)$  is closer to  $g^2(\mathbf{X}^2)$  than  $g^1(\mathbf{X}^1)$ . In order to minimize the upper bound of (36), the order of Bernoullis  $f^{3,1}(\mathbf{X}^1)$  and  $f^{3,2}(\mathbf{X}^2)$  should be swapped. After reordering, we can obtain our new approximating Bernoulli  $g^1(\mathbf{X}^1)$  by merging  $f^{1,1}(\mathbf{X}^1)$ ,  $f^{2,1}(\mathbf{X}^1)$  and  $f^{3,2}(\mathbf{X}^2)$ , and  $g^2(\mathbf{X}^2)$  by merging  $f^{1,2}(\mathbf{X}^2)$ ,  $f^{2,2}(\mathbf{X}^2)$  and  $f^{3,1}(\mathbf{X}^1)$ .

For the EAFS-VMB, in order to minimize the upper bound of (36), a proportion of the weights of assigning  $f^{3,1}(\mathbf{X}^1)$  to  $g^1(\mathbf{X}^1)$  should be shifted to  $g^2(\mathbf{X}^2)$ , and a proportion of the weights of assigning  $f^{3,2}(\mathbf{X}^2)$  to  $g^2(\mathbf{X}^2)$  should be shifted to  $g^1(\mathbf{X}^1)$ . For example, the approximating Bernoulli  $g^1(\mathbf{X}^1)$  can be expressed as a Bernoulli mixture with component  $f^{1,1}(\mathbf{X}^1)$ ,  $f^{2,1}(\mathbf{X}^1)$ ,  $f^{3,1}(\mathbf{X}^1)$  and  $f^{3,2}(\mathbf{X}^2)$ , in which  $f^{3,1}(\mathbf{X}^1)$  has weight 0.05 and  $f^{3,2}(\mathbf{X}^2)$  has weight 0.2; the approximating Bernoulli  $g^2(\mathbf{X}^2)$  can be expressed as the same Bernoulli mixture, but in which  $f^{3,1}(\mathbf{X}^1)$  has weight 0.2 and  $f^{3,2}(\mathbf{X}^2)$  has weight 0.05.

# B. New track formation

We present a greedy method to form new tracks in a reasonable and efficient way. The pseudo code of this merging approach is given in Table II. The intuition behind this proposed method is that we want to merge highly dependent Bernoullis across different MBs so that similar new tracks will not be formed in the same local region and new tracks with significantly different Bernoulli densities will not be merged. In this merging approach, we only merge Bernoullis in different MBs that are considered similar enough. For any

pair of Bernoulli densities, the symmetrized KL divergence, defined as  $D_{\rm SKL}(p||q) = D_{\rm KL}(p||q) + D_{\rm KL}(q||p)$ , is used to measure the similarity.

To start with, the MB densities are sorted in the descending order of their weights, so that  $f^1(\mathbf{X})$  has the highest weight. Note that reordering the MBs will not change the MBM density. Then, we cluster Bernoullis that satisfy the merging criteria into the same group; as a result, the number of new tracks formed is equal to the number of Bernoulli clusters. Lastly, new tracks are formed by only merging Bernoullis within the same group. For the  $\iota$ th ( $\iota \in \hat{\mathbb{I}} \setminus \mathbb{I}$ ) new track in the approximating MB  $g(\mathbf{X})$ , its Bernoulli density can be expressed as

$$\underset{g^{\iota}}{\operatorname{arg\,min}} \ D_{\mathrm{KL}} \bigg( \sum_{j \in \mathbb{J}} W^{j} \\ \times \sum_{i \in \mathbb{J}^{j} \setminus \mathbb{T}} 1_{\mathbb{L}^{\iota - |\mathbb{I}|}} \big( \{ (j, i) \} \big) f^{j, i}(\mathbf{X}^{i}) \bigg| \bigg| g^{\iota}(\mathbf{X}^{\iota}) \bigg), \quad (44)$$

where set  $\mathbb{L}^{\iota-|\mathbb{I}|}$  contains indices of Bernoullis that are within the same group indexed by  $\iota-|\mathbb{I}|$ . Empirical results show that the number of tracks formed using this approach can be kept to a relatively small number. We illustrate this with the following example.

**Example 4.** Consider the same scenario illustrated in Fig. 1. There are five new tracks created in total in the three global hypotheses. If we assume that Bernoulli densities of these five new tracks are mutually independent, it is likely that none of these tracks are detected by the estimator due to their low existence probabilities.

Following the proposed merging approach, we start by merging  $f^{1,3}(\mathbf{X}^3)$  with  $f^{2,3}(\mathbf{X}^3)$  and  $f^{3,3}(\mathbf{X}^3)$ , due to the small symmetrized KL divergence between the pair  $f^{1,3}(\mathbf{X}^3)$ ,  $f^{2,3}(\mathbf{X}^3)$ , and between the pair  $f^{1,3}(\mathbf{X}^3)$ ,  $f^{3,3}(\mathbf{X}^3)$ . Next, we merge  $f^{1,4}(\mathbf{X}^4)$  with  $f^{2,4}(\mathbf{X}^4)$ , due to their high similarity in the sense of symmetrized KL divergence. After merging, the number of Bernoullis in the approximating MB reduces from five to two:  $g^3(\mathbf{X}^3)$  and  $g^4(\mathbf{X}^4)$ :  $g^3(\mathbf{X}^3)$  has existence probability  $\hat{r}_3 = 1$ , and  $g^4(\mathbf{X}^4)$  has existence probability  $\hat{r}_4 = 0.75$ .

#### VIII. GGIW IMPLEMENTATION

Solving the multiple extended target filtering problem requires not only an MTT framework, but also a single extended target model. For the modeling of the spatial distribution, two popular models are the Random Hyper-surface Models [45], [46] and the Random Matrix, also known as Gaussian inverse Wishart (GIW), approach [47], [48]. The former is designed for general star-convex shape; the latter relies on the elliptic shape and it models the spatial distribution of target-generated measurements as Gaussian with unknown mean and covariance. The Gamma GIW (GGIW) model [26] is an extension of the GIW model that incorporates the estimation of Poisson target measurement rate [39].

In this section, some implementation details of the PMB filter are presented. The GGIW implementations of the PMBM

filter and the LMB filter can be, respectively, found in [7] and [12]. To make the comparison easy, we choose to use the GGIW model. In addition, we present strategies regarding how to address the third challenge in an MBM approximation that we outlined in Section V-C, i.e., the merging of a selection of Bernoullis, using a GGIW model.

### A. Single target models

In the GGIW model, it is assumed that target measurements are Gaussian distributed around the target centroid. The extended target state  $\mathbf{x}_k$  is the combination of a Poisson rate  $\gamma_k$  modeling the average number of measurements generated by the target, a random vector  $\xi_k$  describing the target kinematic state, and a random matrix  $\chi_k$  describing the target size and shape, i.e.,  $\mathbf{x}_k = \{\xi_k, \chi_k, \gamma_k\}$ .

The motion models are given by

$$\xi_{k+1} = F(\xi_k) + w_k, \tag{45a}$$

$$\chi_{k+1} = M(\xi_k) \chi_k M(\xi_k)^T, \tag{45b}$$

$$\gamma_{k+1} = \gamma_k, \tag{45c}$$

where  $F(\cdot)$  is a motion model,  $w_k$  is a zero mean Gaussian noise and  $M(\cdot)$  is a transformation matrix.

The measurement likelihood for a single measurement z is

$$\phi(\mathbf{z}_k|\mathbf{x}_k) = \mathcal{N}(\mathbf{z}_k; H_k \xi_k, \chi_k + R_k), \tag{46}$$

where  $H_k$  is the measurement model, and  $R_k$  is the covariance of a zero mean Gaussian noise. The single target conjugate prior for the PPP model (3) with single measurement likelihood (46) is a product of Gamma, Gaussian and inverse Wishart distributions [48]

$$f(\mathbf{x}) = \mathcal{GAM}(\gamma; a, b) \mathcal{N}(\xi; m, P)$$

$$\times \mathcal{IW}_d(\chi; v, V) \triangleq \mathcal{GGIW}(\mathbf{x}; \zeta), \quad (47)$$

where  $\zeta = \{a, b, m, P, v, V\}$  is the set of GGIW density parameters. If we have a PPP birth with GGIW density, then the undetected PPP will have GGIW density, as well as all the Bernoulli components [7].

# B. MBM merging

The GGIW implementations regarding the prediction and update of PPP and Bernoulli components are not presented due to page constraints. The reader is referred to [6], [7] for more details. In this subsection, we present the GGIW implementations regarding the block coordinate descent used to merge the MBM representing existing tracks.

1) E-step: In order to solve the optimization problems (41) and (42) of the E-step, the cross entropy between two Bernoulli-GGIW distributions needs to be calculated. Because the Gamma distributions, the Gaussian distributions and the inverse Wishart distributions are mutually independent, a tractable solution can be analytically derived [39], [40]. See Appendix C for details.

2) M-step: Given a Bernoulli-GGIW mixture, the existence probability of the approximating Bernoulli is a weighted sum of the existence probabilities of each Bernoulli. Suppose that we have a number of Bernoullis indexed by  $n \in \mathbb{N}$ , each of which has existence probability  $r^n$  and GGIW density  $\mathcal{GGIW}(\mathbf{x}^n;\zeta^n)$ . The existence probability of the approximating Bernoulli can be expressed as

$$\hat{r} = \sum_{n \in \mathbb{N}} w^n r^n,\tag{48}$$

where  $w^n$  is the weight of the nth Bernoulli.

The mixture reduction for multivariate Gaussian distribution that minimizes the KL-divergence can be achieved using moment matching. Theorems describing how a sum of an arbitrary number of Gamma components or inverse Wishart components can be merged into a single Gamma or inverse Wishart component are presented in [39] and [40], respectively; they are both performed via analytical minimization of the KL divergence. The same merging techniques also apply to merging the MBM representing new tracks (44). The existence-conditioned GGIW density of the approximating Bernoulli can be obtained by

$$\underset{\mathcal{GGIW}(\hat{\mathbf{x}};\hat{\zeta})}{\arg\min} D_{\mathrm{KL}} \left( \sum_{n \in \mathbb{N}} w^n \mathcal{GGIW}(\mathbf{x}^n; \zeta^n) \middle| \middle| \mathcal{GGIW}(\hat{\mathbf{x}}; \hat{\zeta}) \right). \tag{49}$$

Empirically, we have found that in extended target filtering with GGIW implementation it is generally not advisable to merge all the GGIW components. The main reason is that merging two densities with significantly different extent estimates will result in an approximate density in which the extent estimates are distorted. This problem is exacerbated in the extended PMB filter since the distorted extent states contained by the approximating single MB can easily lead to poor target state estimations in subsequent time steps.

A simple strategy to handle this problem is to use a criterion for deciding which components should be merged. In this work, the KL divergence is used as the similarity measure between any pair of GGIW distributions. The component with the highest weight  $\mathcal{GGIW}(\mathbf{x}^{n^*};\zeta^{n^*})$  is chosen as the comparison baseline, which is merged with all other components  $\mathcal{GGIW}(\mathbf{x}^n;\zeta^n)$  for which it holds

$$D_{\mathrm{KL}}(\mathcal{GGIW}(\mathbf{x}^{n^*}; \zeta^{n^*}) || \mathcal{GGIW}(\mathbf{x}^n; \zeta^n)) < \tau_q, \tag{50}$$

where threshold  $\tau_g$  determines which GGIW components are going to be merged. In this case, the existence-conditioned PDF of the approximating MB can be obtained by

$$\underset{\mathcal{GGIW}(\hat{\mathbf{x}};\hat{\zeta})}{\arg\min} D_{\mathrm{KL}} \left( \sum_{n \in \mathbb{N}: (50)} w^n \mathcal{GGIW}(\mathbf{x}^n; \zeta^n) \middle| \middle| \mathcal{GGIW}(\hat{\mathbf{x}}; \hat{\zeta}) \right). \tag{51}$$

# IX. SIMULATIONS AND RESULTS

In this section we show Monte Carlo (MC) simulation results that compare six different extended target filters:

- 1) LMB filter [12],
- 2) PMBM filter with MBM reduction [7],

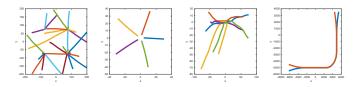


Fig. 2. True target trajectories of four scenarios, from left to right: 1) 27 targets, 2) dense birth, 3) merge/split and 4) nonlinear maneuver.

- 3) TO-PMB<sup>7</sup> filter,
- TO-PMB filter with greedy new tracks merging, denoted as TON-PMB<sup>7</sup>,
- 5) PMB filter using EAFS-VMB, denoted as EAFS-PMB<sup>7</sup>,
- 6) PMB filter using MLA-VMB, denoted as MLA-PMB<sup>7</sup>, in four different simulated scenarios.

# A. State space model

Target motion follows a nearly constant velocity model. A two-dimensional Cartesian coordinate system is used to define measurement and target kinematic parameters. The kinematic state is  $\xi_k = [p_k, v_k]^T$ , describing the target's position  $p_k = [p_{x,k}, p_{y,k}]$  and velocity  $v_k = [v_{x,k}, v_{y,k}]$ . The single measurement is  $\mathbf{z}_k = [z_{x,k}, z_{y,k}]^T$ , where  $z_{x,k}$  and  $z_{y,k}$  describe the position of the measurement. The motion model  $F(\cdot)$  and process noise  $Q_k$  are expressed as

$$F(\xi_k) = \mathbf{I}_2 \otimes \begin{bmatrix} 1 & T \\ 0 & 1 \end{bmatrix} \xi_k, \quad Q_k = \sigma_v^2 \mathbf{I}_2 \otimes \begin{bmatrix} T^4/4 & T^3/2 \\ T^3/2 & T^2 \end{bmatrix},$$

where T=1s is the sampling period, and  $\sigma_v$  is the standard deviation of velocity noise. The random matrix  $V_k$  in the inverse Wishart distribution is two-dimensional. Because the kinematic state motion model is constant velocity, the extent transformation function M is an identity matrix, i.e.,  $M(\xi_k) = I_2$ .

# B. Performance evaluation

For GGIW-PMB and GGIW-LMB, the target states are extracted by taking the mean vector of all Bernoullis with existence probability larger than 0.5. For GGIW-PMBM, target state extraction is performed analogously, but only from the MB with the highest weight.

For performance evaluation of extended object estimates with ellipsoidal extents, a comparison study has shown that a good choice is the Gaussian Wasserstein Distance (GWD) metric [49]. To evaluate the performance of different multi-target filtering algorithms, we use both the Optimal Sub-pattern Assignment (OSPA) metric [50] and the Generalized OSPA (GOSPA) metric [51] with parameters  $\alpha=2, c=10, p=1$ ; both metrics are integrated with GWD as the base distance measure. Compared to OSPA, GOSPA is not normalized by the cardinality of the largest set and it penalizes cardinality errors differently [51], which allows for the decomposition of the estimation error into three different categories: localization error, missed detection error and false detection error. See [7, Eq. (32), (33)] for explicit mathematical expressions.

 $^7\mathrm{MATLAB}$  code of different variants of extended target PMB filter is available at https://github.com/yuhsuansia/Extended-target-PMB-filter.

# C. Simulation study

We evaluate the filters in four different scenarios. True target trajectories are shown in Fig. 2. In the first scenario, 27 randomly generated targets are born from four localized positions, and they appear in and disappear from the surveillance area at different time steps. The parameters were set to  $p^D = 0.90, p^S = 0.99, \lambda = 60$  and  $\gamma \in \{7, 8, 9\}.$ This scenario illustrates how the different filters behave with a high target number and high clutter density scenario. In the second scenario, five targets are born at a very short distance from each other at the same time step. The parameters were set to  $p^D = 0.90, p^S = 0.99, \lambda = 20$  and  $\gamma = 10$ . This scenario tests different filters capabilities of handling a dense birth. In the third scenario, five targets first get close to each other and then separate. The parameters were set to  $p^D = 0.7, p^S = 0.99, \lambda = 10$  and  $\gamma = 5$ . This scenario tests different filters capabilities of handling coalescence under low detection probability. In the fourth scenario, two targets first get close, and then they maneuver in close proximity before splitting; the data association problem is very challenging in this scenario due to the coalescence and the highly-nonlinear motion when targets are turning. The parameters were set to  $p^D = 0.98, p^S = 0.99, \lambda = 10 \text{ and } \gamma \in \{10, 20\}.$ 

For solving the data association problem, we first apply DBSCAN [52] with different distance thresholds between 0.1 and 5 to obtain a set of measurement partitions. For each unique measurement partition, we then use Murty's algorithm to find the  $\lceil 20 \cdot W_A \rceil$  ( $W_A = 1$  for PMB and LMB) best cell-to-track assignments; these are pruned to only contain the MBs that correspond to 99.99% of the likelihood. For all the compared filters, Bernoullis with existence probability smaller than 0.001 are pruned. For PMBM, the merging threshold in MBM reduction is set to 0.1. For EAFS-PMB and MLA-PMB, the E-step and the M-step in VMB algorithm are iterated until the decrease of the cross entropy from one iteration to next is smaller than 0.001.

For each scenario, the result is averaged over 100 MC trials. The filtering performance of different filters in terms of OSPA/GOSPA error and cycle time<sup>8</sup> are shown in Table III, the GOSPA performance over time is shown in Fig. 3<sup>9</sup>, and an analysis regarding the convergence of two VMB implementations is given in Table IV<sup>10</sup>. It can be seen from the results that, in general, the PMBM filter achieves the lowest estimation error, with PMB filter the second, and the LMB filter has the highest estimation error, in terms of both OSPA and GOSPA errors. From the perspective of average cycle time per MC run, the PMB filter has the lowest computational cost. In the forth scenario, none of the filters exhibit good estimation performance due to the motion model mismatch in the prediction step, and the opposite results of OSPA and GOSPA is due to the fact that they penalize the cardinality errors differently.

 $<sup>^8</sup>$ MATLAB implementation on Intel(R) Core(TM) i7-7700K @ 4.20GHz.

<sup>&</sup>lt;sup>9</sup>To not clutter the figure with too many curves, only the results of LMB, PMBM and MLA-PMB are presented.

<sup>&</sup>lt;sup>10</sup>This table only presents results for time steps when VMB algorithm was actually implemented. In some cases, if there is only one MB has dominant weight, then there is no need to perform any merging.

TABLE III

SIMULATION RESULTS: THE SUM OF ESTIMATION ERRORS AND CYCLE TIME (SECONDS), AVERAGED OVER MC RUNS.

LEGEND: O-OSPA; GO-GOSPA; LE-LOCATION ERROR; NF-NUMBER OF FALSE DETECTION; NM-NUMBER OF MISSED DETECTION; T-CYCLE TIME

	Scenario 1: 27 targets							Scenario 2: dense birth						Scenario 3: merge/split						Scenario 4: nonlinear maneuver					
Filter	0	GO	LE	NF	NM	T	0	GO	LE	NF	NM	T	0	GO	LE	NF	NM	T	О	GO	LE	NF	NM	T	
PMBM	95.1	721.5	567.7	15.5	15.2	74.8	37.7	125.8	46.8	2.9	12.9	15.0	155.1	550.8	225.0	13.5	51.7	81.6	2682.0	3590.7	385.3	143.4	497.7	16.0	
TO-PMB	97.6	746.3	568.5	15.6	19.9	20.8	41.3	153.2	58.5	6.3	12.6	2.0	177.3	664.3	245.1	30.5	53.3	8.7	2699.4	3565.5	377.7	136.6	503.0	14.4	
TON-PMB	96.1	735.9	569.9	16.7	16.5	19.2	41.1	150.9	58.8	5.4	13.0	1.8	169.3	609.7	227.4	21.7	54.7	6.5	2699.4	3570.8	367.5	137.6	503.0	14.6	
MLA-PMB	95.5	732.5	569.6	16.2	16.4	18.1	39.5	145.0	58.9	4.9	12.3	1.9	167.4	597.9	216.1	20.9	55.4	7.9	2699.5	3561.9	367.2	135.9	503.1	14.5	
EAFS-PMB	95.5	735.0	572.3	16.2	16.4	18.9	39.6	144.7	58.9	4.9	12.3	1.9	167.7	600.0	214.1	21.4	55.8	6.8	2699.5	3561.9	367.2	135.9	503.1	14.5	
LMB	135.5	1069.9	923.5	10.2	19.1	28.4	62.6	259.5	84.4	12.4	22.7	3.0	192.7	741.5	294.9	26.0	63.3	14.3	2591.9	3982.5	421.9	235.9	476.2	25.3	

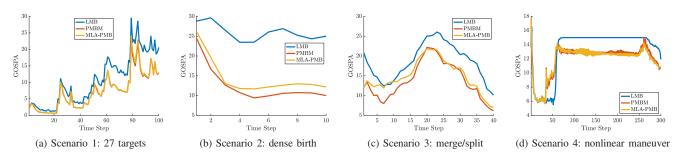


Fig. 3. Simulation results comparison of LMB, PMBM and MLA-PMB in terms of GOSPA.

#### TABLE IV

CONVERGENCE ANALYSIS OF VMB ALGORITHM, AVERAGED OVER MC RUNS AND TIME STEPS WHEN VMB WAS ACTUALLY IMPLEMENTED.

LEGEND:NI-NUMBER OF ITERATIONS TILL CONVERGENCE; NT-NUMBER OF TIME STEPS BEING IMPLEMENTED; CEBV-CROSS ENTROPY BETWEEN MBM

AND APPROXIMATING MB BEFORE APPLYING VMB; CEAV-CROSS ENTROPY BETWEEN MBM AND APPROXIMATING MB AFTER APPLYING VMB;

D-DIFFERENCE BETWEEN CEBV AND CEAV

	Scenario 1: 27 targets (100 steps)						Scenario 2: dense birth (10 steps)					Scenario 3: merge/split (40 steps)						Scenario 4: nonlinear maneuver (300 steps)				
Filter	NI	NT	CEBV	CEAV	D	NI	NT	CEBV	CEAV	D	NI	NT	CEBV	CEAV	D	NI	NT	CEBV	CEAV	D		
EAFS-PMB	2.22	43.04	7.68	7.59	0.10	3.67	7.91	26.82	26.27	0.30	3.02	36.51	19.60	19.05	0.54	1.09	5.42	16.31	16.27	0.04		
MLA-PMB	2.22	43.00	7.77	7.63	0.14	3.67	7.91	26.80	26.28	0.29	2.99	36.20	19.92	19.37	0.55	1.09	5.42	16.31	16.28	0.04		

By employing the proposed greedy method for merging new tracks, TON-PMB shows less estimation error than TO-PMB in terms of GOSPA/OSPA; the difference is most noticeable in the results of the first three scenarios. The PMB filters using variational approximation have better estimation performance than their counterparts without variational approximation, especially in the scenarios with coalescence. By comparing the cycle time of different variants of PMB, we can find that the additional computational cost brought by VMB and the greedy new tracks merging method is small because the cycle time is dominated by solving the data association problem.

The fast convergency of VMB can be verified from Table IV that, in most cases, both EAFS-VMB and MLA-VMB can converge in four iterations. Table IV also lists the numerical values of the cross entropy between the MBM and the approximating MB before and after applying VMB algorithms; they are denoted as CEAV and CEBV, respectively. The former is calculated using the track-oriented merging solution of (35), and the latter is calculated based on two different approximate solutions (41), (42) of (35). Because both EAFS-VMB and MLA-VMB are initialized with the track-oriented merging, the difference between CEAV and CEBV indicates if and how well the variational approximation method works. These results demonstrate that approximate solutions obtained using either EAFS-VMB (41) or MLA-VMB (42) can yield a lower KL divergence compared to the track-oriented merging approach.

The two variants of VMB implementations present similar estimation performance, in terms of both OSPA and GOSPA

error, in all the compared scenarios. The average cycle time of MLA-VMB is less than it of EAFS-PMB in the first scenario, but the corresponding result becomes opposite in the third scenario. It should be noted, however, that the average computational time per VMB iteration depends on the optimization solver being used to solve (41) or (42). To conclude, the PMB filters using variational approximation achieves an appealing trade-off between computational time and estimation performance.

# X. CONCLUSIONS

This paper presents an efficient multiple extended target filtering algorithm based on an approximation of a PMBM posterior density as a PMB, along with its GGIW implementation. A thorough simulation study shows that the presented extended target PMB filter can inherit the good performance of the extended target PMBM filter but with lower computational complexity.

#### APPENDIX A

In this appendix, we show how to merge a mixture of Bernoulli densities, in the sense of minimizing the KL divergence.

Let  $f^{\mathbb{H}}(\mathbf{X}) = \sum_{h \in \mathbb{H}} w^h f^h(\mathbf{X})$  be a mixture of Bernoulli densities  $f^h(\mathbf{X})$ , where the existence-conditioned PDF of  $f^h(\mathbf{X})$  is from a family of distributions  $\mathcal{F}$ , i.e.,  $f^h(\mathbf{x}) \in \mathcal{F}$ ,  $\forall h \in \mathbb{H}$ . To approximate the Bernoulli mixture  $f^{\mathbb{H}}(\mathbf{X})$  by a single Bernoulli density  $\hat{f}(\mathbf{X})$ , whose existence-conditioned

PDF is from the the same family of distributions, i.e.,  $\hat{f}(\mathbf{x}) \in \mathcal{F}$ , the approximating Bernoulli density  $\hat{f}(\mathbf{X})$  that minimizes the KL divergence

$$D_{KL}(f^{\mathbb{H}}(\mathbf{X})||\hat{f}(\mathbf{X})) = \int f^{\mathbb{H}}(\mathbf{X}) \log \left(\frac{f^{\mathbb{H}}(\mathbf{X})}{\hat{f}(\mathbf{X})}\right) \delta \mathbf{X}, \quad (52)$$

has parameters [8]:

$$\hat{r} = \sum_{h \in \mathbb{H}} w^h r^h, \tag{53a}$$

$$\hat{f}(\mathbf{x}) = \underset{f \in \mathcal{F}}{\arg \min} \ D_{KL} \left( \sum_{h \in \mathbb{H}} w^h r^h f^h(\mathbf{x}) \middle| \middle| f(\mathbf{x}) \right). \tag{53b}$$

For distributions from the exponential family, the KL divergence minimization (53b) can be analytically solved by matching the expected sufficient statistics, see, e.g., [53, Section 10.7], [54].

#### APPENDIX B

In this appendix, we prove Theorem 1.

*Proof.* The problem of (33) can be reformulated as

$$\underset{g}{\operatorname{arg\,min}} - \int f(\mathbf{X}) \log g(\mathbf{X}) \delta \mathbf{X}, \tag{54}$$

which can be further rewritten as

$$\underset{g^{1},g^{2}}{\operatorname{arg\,min}} - \sum_{j \in \mathbb{J}} W^{j} \int \sum_{\mathbf{X}^{1} \uplus \mathbf{X}^{2} = \mathbf{X}} f^{j}(\mathbf{X}^{1}) f^{j}(\mathbf{X}^{2})$$

$$\times \log \left( \sum_{\hat{\mathbf{X}}^{1} \uplus \hat{\mathbf{X}}^{2} = \mathbf{X}} g^{1}(\hat{\mathbf{X}}^{1}) g^{2}(\hat{\mathbf{X}}^{2}) \right) \delta \mathbf{X}. \quad (55)$$

According to [9, Thm. 5], the multi-target set integral can be decomposed into a series of set integrals, we can rewrite the objective function of the minimization problem (55) as

$$J([g^{1}, g^{2}]) = -\sum_{j \in \mathbb{J}} W^{j} \int \int f^{j}(\mathbf{X}^{1}) f^{j}(\mathbf{X}^{2})$$

$$\times \log \left( \sum_{\hat{\mathbf{X}}^{1} \uplus \hat{\mathbf{X}}^{2} = \mathbf{X}^{1} \uplus \mathbf{X}^{2}} g^{1}(\hat{\mathbf{X}}^{1}) g^{2}(\hat{\mathbf{X}}^{2}) \right) \delta \mathbf{X}^{1} \delta \mathbf{X}^{2}. \quad (56)$$

Applying the log-sum inequality, an upper bound can be obtained as [9, Sec. III.A]

$$J([g^{1}, g^{2}]) \leq -\sum_{j \in \mathbb{J}} W_{j} \iint f^{j}(\mathbf{X}^{1}) f^{j}(\mathbf{X}^{2})$$

$$\times \log (g^{1}(\mathbf{X}^{1}) g^{2}(\mathbf{X}^{2})) \delta \mathbf{X}^{1} \delta \mathbf{X}^{2}$$

$$= -\sum_{j \in \mathbb{J}} W_{j} \left( \int f^{j}(\mathbf{X}^{1}) \log (g^{1}(\mathbf{X}^{1})) \delta \mathbf{X}^{1} \right)$$

$$+ \int f^{j}(\mathbf{X}^{2}) \log (g^{2}(\mathbf{X}^{2})) \delta \mathbf{X}^{2}$$

$$= -\int \sum_{j \in \mathbb{J}} W_{j} f^{j}(\mathbf{X}^{1}) \log (g^{1}(\mathbf{X}^{1})) \delta \mathbf{X}^{1}$$

$$-\int \sum_{j \in \mathbb{J}} W_{j} f^{j}(\mathbf{X}^{2}) \log (g^{2}(\mathbf{X}^{2})) \delta \mathbf{X}^{2}.$$

$$(57)$$

The objective function in the minimization problem (55) has an upper bound that, when we minimize over  $g^1(\cdot)$  and  $g^2(\cdot)$ , can be broken down into two separate minimization problems

$$\min_{g^{1},g^{2}} \left[ -\int \sum_{j\in\mathbb{J}} \mathcal{W}_{j} f^{j}(\mathbf{X}^{1}) \log \left(g^{1}(\mathbf{X}^{1})\right) \delta \mathbf{X}^{1} \right] 
-\int \sum_{j\in\mathbb{J}} \mathcal{W}_{j} f^{j}(\mathbf{X}^{2}) \log \left(g^{2}(\mathbf{X}^{2})\right) \delta \mathbf{X}^{2} \right] 
= \min_{g^{1}} \left[ -\int \sum_{j\in\mathbb{J}} \mathcal{W}_{j} f^{j}(\mathbf{X}^{1}) \log \left(g^{1}(\mathbf{X}^{1})\right) \delta \mathbf{X}^{1} \right] 
+ \min_{g^{2}} \left[ -\int \sum_{j\in\mathbb{J}} \mathcal{W}_{j} f^{j}(\mathbf{X}^{2}) \log \left(g^{2}(\mathbf{X}^{2})\right) \delta \mathbf{X}^{2} \right]$$
(61)

Note that the arguments that minimize these two objective functions are the same as the arguments that minimize the KL divergences,

$$\underset{g^1}{\operatorname{arg\,min}} D\bigg(\sum_{j\in\mathbb{T}} \mathcal{W}_j f^j(\mathbf{X}^1) \Big| \Big| g^1(\hat{\mathbf{X}}^1)\bigg), \tag{62}$$

$$\underset{g^2}{\operatorname{arg\,min}} D\bigg(\sum_{j\in\mathbb{J}} \mathcal{W}_j f^j(\mathbf{X}^2) \bigg| \bigg| g^2(\hat{\mathbf{X}}^2)\bigg). \tag{63}$$

This proves Theorem 1.

#### APPENDIX C

In this appendix, we show how to calculate the cross Entropy between two Bernoulli-GGIWs. Suppose  $f^h(\mathbf{X})$  and  $g^i(\mathbf{X})$  are two Bernoulli processes with the following form

$$f^{h}(\mathbf{X}) = \begin{cases} 1 - r^{h}, & \mathbf{X} = \emptyset \\ r^{h} \mathcal{GGIW}(\mathbf{x}^{h}; \zeta^{h}), & \mathbf{X} = \{\mathbf{x}\} \end{cases}, \tag{64a}$$

$$g^{i}(\mathbf{X}) = \begin{cases} 1 - r^{i}, & \mathbf{X} = \emptyset, \\ r^{i} \mathcal{G} \mathcal{G} \mathcal{I} \mathcal{W}(\mathbf{x}^{i}; \zeta^{i}), & \mathbf{X} = \{\mathbf{x}\} \end{cases}, \tag{64b}$$

where  $\zeta = \{a, b, m, P, v, V\}$  is the set of GGIW density parameters. Because the Gaussian, Gamma and inverse Wishart distributions are mutually independent, the cross entropy between  $f^h(\mathbf{X})$  and  $g^i(\mathbf{X})$  can be expressed in closed form as:

$$-\int f^{h}(\mathbf{X}) \log g^{i}(\mathbf{X}) \delta \mathbf{X} = -(1 - r^{h}) \log(1 - r^{i})$$

$$-r^{h} \log r^{i} - r^{h} \left( \int \mathcal{N}(\xi^{h}; m^{h}, P^{h}) \log \mathcal{N}(\xi^{i}; m^{i}, \hat{P}^{i}) d\xi \right)$$

$$+ \int \mathcal{GAM}(\gamma^{h}; a^{h}, b^{h}) \log \mathcal{GAM}(\gamma^{i}; a^{i}, b^{i}) d\gamma$$

$$+ \int \mathcal{IW}(\chi^{h}; v^{h}, V^{h}) \log \mathcal{IW}(\chi^{i}; v^{i}, V^{i}) d\chi , \quad (65)$$

where

$$\int \mathcal{N}(\xi^h; m^h, P^h) \log \mathcal{N}(\xi^i; m^i, P^i) d\xi =$$

$$-\frac{d}{2} \log(2\pi) - \frac{1}{2} \log(\det(P^i))$$

$$-\frac{1}{2} \text{Tr} \left( \left( P^h + (m^h - m^i)(m^h - m^i)^T \right) \left( P^i \right)^{-1} \right), \quad (66a)$$

$$\int \mathcal{GAM}(\gamma^h; a^h, b^h) \log \mathcal{GAM}(\gamma^i; a^i, b^i) d\gamma = a^i \log b^i$$
$$-\log \Gamma(a^i) + (a^i - 1)(\psi_0(a^h) - \log b^h) - b^i \frac{a^h}{b^h}, \quad (66b)$$

and

$$\int \mathcal{IW}(\chi^h; v^h, V^h) \log \mathcal{IW}(\chi^i; v^i, V^i) d\chi =$$

$$-\frac{(v^i - d - 1)d}{2} \log 2 + \frac{v^i - d - 1}{2} \log(\det(V^i))$$

$$-\log \Gamma_d \left(\frac{v^i - d - 1}{2}\right) - \frac{v^i}{2} \left(\log(\det(V^h)) - d\log 2\right)$$

$$-\sum_{j=1}^d \psi_0 \left(\frac{v^h - d - j}{2}\right) - \frac{1}{2} \text{Tr}\left((v^h - d - 1)(V^h)^{-1}V^i\right).$$
(66c)

#### REFERENCES

- B.-n. Vo, M. Mallick, Y. Bar-shalom, S. Coraluppi, R. Osborne III, R. Mahler, and B.-t. Vo, "Multitarget tracking," Wiley Encyclopedia of Electrical and Electronics Engineering, pp. 1–15, 1999.
- [2] R. P. Mahler, Advances in Statistical Multisource-Multitarget Information Fusion. Artech House Norwood, MA, 2014.
- [3] B.-T. Vo and B.-N. Vo, "Labeled random finite sets and multi-object conjugate priors," *IEEE Transactions on Signal Processing*, vol. 61, no. 13, pp. 3460–3475, 2013.
- [4] A. F. García-Fernández, J. Williams, K. Granström, and L. Svensson, "Poisson multi-Bernoulli mixture filter: direct derivation and implementation," *IEEE Transactions on Aerospace and Electronic Systems*, 2018.
- [5] Y. Xia, K. Granström, L. Svensson, and Á. F. García-Fernández, "Performance evaluation of multi-Bernoulli conjugate priors for multi-target filtering," in *Proceedings of International Conference on Information Fusion*. IEEE, 2017.
- [6] K. Granström, M. Fatemi, and L. Svensson, "Gamma Gaussian inverse-Wishart poisson multi-Bernoulli filter for extended target tracking," in *Proceedings of International Conference on Information Fusion*. IEEE, 2016, pp. 893–900.
- [7] —, "Poisson multi-Bernoulli conjugate prior for multiple extended object filtering," *IEEE Transactions on Aerospace and Electronic Systems*, 2019.
- [8] J. L. Williams, "Marginal multi-Bernoulli filters: RFS derivation of MHT, JIPDA, and association-based member," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 51, no. 3, pp. 1664–1687, 2015.
- [9] —, "An efficient, variational approximation of the best fitting multi-Bernoulli filter," *IEEE Transactions on Signal Processing*, vol. 63, no. 1, pp. 258–273, 2015.
- [10] K. Granström, L. Svensson, S. Reuter, Y. Xia, and M. Fatemi, "Likelihood-based data association for extended object tracking using sampling methods," *IEEE Transactions on Intelligent Vehicles*, vol. 3, no. 1, pp. 30–45, 2018.
- [11] L. Svensson, D. Svensson, M. Guerriero, and P. Willett, "Set JPDA filter for multitarget tracking," *IEEE Transactions on Signal Processing*, vol. 59, no. 10, pp. 4677–4691, 2011.
- [12] M. Beard, S. Reuter, K. Granström, B.-T. Vo, B.-N. Vo, and A. Scheel, "Multiple extended target tracking with labeled random finite sets," *IEEE Transactions on Signal Processing*, vol. 64, no. 7, pp. 1638–1653, 2016.
- [13] M. Schuster, J. Reuter, and G. Wanielik, "Probabilistic data association for tracking extended targets under clutter using random matrices," in *Proceedings of International Conference on Information Fusion*. IEEE, 2015, pp. 961–968.
- [14] R. Streit, "JPDA intensity filter for tracking multiple extended objects in clutter," in *Proceedings of International Conference on Information Fusion*. IEEE, 2016, pp. 1477–1484.
- [15] G. Vivone and P. Braca, "Joint probabilistic data association tracker for extended target tracking applied to x-band marine radar data," *IEEE Journal of Oceanic Engineering*, vol. 41, no. 4, pp. 1007–1019, 2016.
- [16] M. Wieneke and W. Koch, "Probabilistic tracking of multiple extended targets using random matrices," in *Signal and Data Processing of Small Targets* 2010, vol. 7698. International Society for Optics and Photonics, 2010, p. 769812.

- [17] M. Wieneke and S. J. Davey, "Histogram pmht with target extent estimates based on random matrices," in *Proceedings of International Conference on Information Fusion*. IEEE, 2011, pp. 1–8.
- [18] M. Wieneke and W. Koch, "A pmht approach for extended objects and object groups," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 48, no. 3, pp. 2349–2370, 2012.
- [19] K. Granström, M. Baum, and S. Reuter, "Extended object tracking: Introduction, overview and applications," *Journal of Advances in Information Fusion*, vol. 12, no. 2, 2017.
- [20] R. P. Mahler, "Multitarget bayes filtering via first-order multitarget moments," *IEEE Transactions on Aerospace and Electronic systems*, vol. 39, no. 4, pp. 1152–1178, 2003.
- [21] K. Granström, C. Lundquist, and O. Orguner, "Extended target tracking using a Gaussian-mixture PHD filter," *IEEE Transactions on Aerospace* and Electronic Systems, vol. 48, no. 4, pp. 3268–3286, 2012.
- [22] K. Granström and U. Orguner, "A PHD filter for tracking multiple extended targets using random matrices," *IEEE Transactions on Signal Processing*, vol. 60, no. 11, pp. 5657–5671, 2012.
- [23] A. Swain and D. Clark, "Extended object filtering using spatial independent cluster processes," in *Proceedings of International Conference on Information Fusion (FUSION)*. IEEE, 2010, pp. 1–8.
- [24] —, "The PHD filter for extended target tracking with estimable extent shape parameters of varying size," in *Proceedings of International Conference on Information Fusion (FUSION)*. IEEE, 2012, pp. 1111–1118.
- [25] R. Mahler, "PHD filters of higher order in target number," IEEE Transactions on Aerospace and Electronic Systems, vol. 43, no. 4, 2007.
- [26] C. Lundquist, K. Granström, and U. Orguner, "An extended target CPHD filter and a gamma Gaussian inverse Wishart implementation," *IEEE Journal of Selected Topics in Signal Processing*, vol. 7, no. 3, pp. 472–483, 2013.
- [27] G. Zhang, F. Lian, and C. Han, "CBMeMBer filters for nonstandard targets, I: extended targets," in *Proceedings of International Conference* on *Information Fusion*. IEEE, 2014, pp. 1–6.
- [28] K. Granström, L. Svensson, Y. Xia, J. Williams, and Á. F. García-Femández, "Poisson multi-Bernoulli mixture trackers: continuity through random finite sets of trajectories," in *Proceedings of International Conference on Information Fusion*. IEEE, 2018, pp. 1–5.
- [29] K. Granström, L. Svensson, Y. Xia, Á. F. Williams, García-Fernández, and J. L. Williams, "Extended target Poisson multi-Bernoulli mixture trackers based on sets of trajectories," in *Proceedings of International Conference on Information Fusion*. IEEE, 2019.
- [30] Á. F. García-Fernández, Y. Xia, K. Granström, L. Svensson, and J. Williams, "Gaussian implementation of the multi-Bernoulli mixture filter," in *Proceedings of International Conference on Information Fu*sion. IEEE, 2019.
- [31] Á. F. García-Fernández, L. Svensson, and M. R. Morelande, "Multiple target tracking based on sets of trajectories," *IEEE Transactions on Aerospace and Electronic Systems*, 2019.
- [32] B.-N. Vo, B.-T. Vo, and D. Phung, "Labeled random finite sets and the bayes multi-target tracking filter," *IEEE Transactions on Signal Processing*, vol. 62, no. 24, pp. 6554–6567, 2014.
- [33] B.-N. Vo, B.-T. Vo, and H. G. Hoang, "An efficient implementation of the generalized labeled multi-Bernoulli filter," *IEEE Transactions on Signal Processing*, vol. 65, no. 8, pp. 1975–1987, 2016.
- [34] M. Beard, S. Reuter, K. Granström, B.-T. Vo, B.-N. Vo, and A. Scheel, "A generalised labelled multi-Bernoulli filter for extended multi-target tracking," in *Proceedings of International Conference on Information Fusion*. IEEE, 2015, pp. 991–998.
- [35] S. Reuter, B.-T. Vo, B.-N. Vo, and K. Dietmayer, "The labeled multi-Bernoulli filter," *IEEE Transactions on Signal Processing*, vol. 62, no. 12, pp. 3246–3260, 2014.
- [36] K. Gilholm and D. Salmond, "Spatial distribution model for tracking extended objects," *IEEE Proceedings-Radar, Sonar and Navigation*, vol. 152, no. 5, pp. 364–371, 2005.
- [37] S. S. Blackman, "Multiple hypothesis tracking for multiple target tracking," *IEEE Aerospace and Electronic Systems Magazine*, vol. 19, no. 1, pp. 5–18, 2004.
- [38] J. L. Williams, "Hybrid Poisson and multi-Bernoulli filters," in *Proceedings of International Conference on Information Fusion*. IEEE, 2012, pp. 1103–1110.
- [39] K. Granström and U. Orguner, "Estimation and maintenance of measurement rates for multiple extended target tracking," in *Proceedings of International Conference on Information Fusion*. IEEE, 2012, pp. 2170–2176.

- [40] ——, "On the reduction of Gaussian inverse Wishart mixtures," in Proceedings of International Conference on Information Fusion. IEEE, 2012, pp. 2162–2169.
- [41] J. Olofsson, C. Veibäck, and G. Hendeby, "Sea ice tracking with a spatially indexed labeled multi-Bernoulli filter," in *Proceedings of International Conference on Information Fusion*. IEEE, 2017, pp. 1–8.
- [42] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *Journal of the royal statistical society. Series B (methodological)*, pp. 1–38, 1977.
- [43] G. B. Dantzig, "Application of the simplex method to a transportation problem," *Activity analysis of production and allocation*, vol. 13, pp. 359–373, 1951.
- [44] H. W. Kuhn, "The Hungarian method for the assignment problem," *Naval research logistics quarterly*, vol. 2, no. 1-2, pp. 83–97, 1955.
- [45] M. Baum and U. D. Hanebeck, "Extended object tracking with random hypersurface models," *IEEE Transactions on Aerospace and Electronic* Systems, vol. 50, no. 1, pp. 149–159, 2014.
- [46] N. Wahlström and E. Özkan, "Extended target tracking using Gaussian processes," *IEEE Transactions on Signal Processing*, vol. 63, no. 16, pp. 4165–4178, 2015.
- [47] J. W. Koch, "Bayesian approach to extended object and cluster tracking using random matrices," *IEEE Transactions on Aerospace and Electronic* Systems, vol. 44, no. 3, 2008.
- [48] M. Feldmann, D. Franken, and W. Koch, "Tracking of extended objects and group targets using random matrices," *IEEE Transactions on Signal Processing*, vol. 59, no. 4, pp. 1409–1420, 2011.
- [49] S. Yang, M. Baum, and K. Granström, "Metrics for performance evaluation of elliptic extended object tracking methods," in *Proceedings* of International Conference on Multisensor Fusion and Integration for Intelligent Systems (MFI). IEEE, 2016, pp. 523–528.
- [50] D. Schuhmacher, B.-T. Vo, and B.-N. Vo, "A consistent metric for performance evaluation of multi-object filters," *IEEE Transactions on Signal Processing*, vol. 56, no. 8, pp. 3447–3457, 2008.
- [51] A. S. Rahmathullah, Á. F. García-Fernández, and L. Svensson, "Generalized optimal sub-pattern assignment metric," in *Proceedings of International Conference on Information Fusion*. IEEE, 2017, pp. 1–8.
- [52] M. Ester, H.-P. Kriegel, J. Sander, X. Xu et al., "A density-based algorithm for discovering clusters in large spatial databases with noise." in Kdd, vol. 96, no. 34, 1996, pp. 226–231.
- [53] C. M. Bishop, Pattern recognition and machine learning. New York, USA: Springer, 2006.
- [54] T. Ardeshiri, K. Granström, E. Özkan, and U. Orguner, "Greedy reduction algorithms for mixtures of exponential family," *IEEE Signal Processing Letters*, vol. 22, no. 6, pp. 676–680, Jun. 2015.