Labeled Random Finite Sets and Multi-Object Conjugate Priors

Ba-Tuong Vo*, Ba-Ngu Vo†

Abstract—**The objective of multi-object estimation is to simultaneously estimate the number of objects and their states from a set of observations in the presence of data association uncertainty, detection uncertainty, false observations and noise. This estimation problem can be formulated in a Bayesian framework by modeling the (hidden) set of states and set of observations as random finite sets (RFSs) that covers thinning, Markov shifts and superposition. A prior for the hidden RFS together with the likelihood of the realization of the observed RFS gives the posterior distribution via the application of Bayes rule. We propose a new class of RFS distributions that is conjugate with respect to the multi-object observation likelihood and closed under the Chapman-Kolmogorov equation. This result is tested via a Bayesian multi-target tracking example.**

Index Terms—**Random finite set, marked point process, conjugate prior, Bayesian estimation, target tracking.**

I. Introduction

Inference in multi-object systems involves the simultaneous estimation of the number of objects along with their states, and lies at the heart of two seemingly unrelated fields–target tracking and spatial statistics. These two fields span a diverse range of application areas from defence [1], computer vision [2], robotics [3], [4] to agriculture/forestry [5], [6], [7] and epidemiology/public health [8]. See [9], [10] for a more comprehensive survey of application areas. In target tracking the objective is to locate an unknown and time varying set of targets from sensor detections [11], [1], while in spatial statistics the objective is to analyze and characterize the underlying distribution of spatial point patterns from observed data [9], [10].

A basic observation model, in both spatial statistics and multi-target tracking, is a random finite set (RFS) that accounts for thinning (of misdetected objects), Markov shifts (of true observations) and superposition (of false observations) [1], [6]. Since the number of objects is not known, the set of objects, called the *multi-object state*, is also modeled as an RFS. A prior distribution on the multi-object state can be combined with the observation likelihood via Bayes rule to give the posterior distribution.

Acknowledgement: This work is supported by the Australian Research Council under the Discovery Early Career Researcher Award DE120102388 and Future Fellowship FT0991854

B.-T. Vo and B.-N. Vo are with the Department of Electrical and Computer Engineering, Curtin University, Bentley, WA 6102, Australia (email: ba-tuong.vo@curtin.edu.au, ba-ngu.vo@curtin.edu.au)

Computing the posterior via Bayes rule requires integrating the product of the prior and likelihood function to obtain the normalizing constant. This integration poses significant practical challenges and is a key research theme in Bayesian methods [12]. On one hand, cheap and powerful computers have motivated the investigation of numerical approximations to these integrals [13]. On the other hand, considerable effort has been dedicated to deriving pairs of likelihood functions and prior distributions with tractable analytic normalizing constants (and other convenient mathematical properties). These families of prior distributions–known as conjugate priors–play an important role in Bayesian inference and were first formalized by Raiffa and Schlaifer [14]. Well known classes of conjugate priors are available in the exponential family [15]. In Bayesian nonparametric inference, the Dirichlet Process is a conjugate prior that is widely use in the field of machine learning [16].

Conjugacy is highly desirable in multi-object inference, since computing the posterior of non-conjugate RFS priors are intractable, due to the curse of dimensionality and the inherently combinatorial nature of the problem. Current numerical integration techniques for approximating the normalizing constant are not suitable for online inference, except for scenarios with a small number of objects. While conjugate priors are available for a special type of likelihood function (separable) [17], such priors are not yet available for the standard multi-object likelihood function that models thinning, Markov shift and superposition.

To be of practical use, a conjugate prior family must also be flexible enough to model our prior degree of belief in the parameters of interest. In applications where the multi-object state stochastically evolves in time, we require a Markov multi-object transition model involving thinning (object death), Markov shift (of surviving objects) and superposition (object birth). Inference then involves recursive estimation of the multi-object state at each time step and the Chapman Kolmogorov equation is needed to perform a prediction step to account for multi-object evolution before Bayes rule can be applied [1]. In general, solutions to the multi-object Chapman-Kolmogorov equation cannot be easily determined. Moreover, even if such a solution were available, it is not necessarily a conjugate prior.

This paper introduces the notion of labeled RFSs, and conjugate priors for the standard multi-object likelihood function, which are also closed under the multi-object Chapman-Kolmogorov equation with respect to the multi-object transition kernel. These results provide analytic solutions to the multi-object inference and filtering problems. In multi-object filtering, if we start with the proposed conjugate initial prior, then all subsequent predicted and posterior distributions have the same form as the initial prior. We illustrate an application of these results by a multi-target tracking example. Preliminary results have been published in [18]. The current work is a more complete study with stronger results. A related, but independent, investigation on conjugate priors for (unlabeled) RFS is also being undertaken, with work in progress reported in [19].

The paper is organized as follows. A brief overview of conjugate priors and multi-object estimation is provided in section II while labeled RFSs and their properties are introduced in Section III. Section IV introduces the generalized labeled multi-Bernoulli RFS, and show that it is conjugate prior as well as being closed under the multi-object Chapman-Kolmogorov equation. Section V establishes similar results on a smaller family within the class of generalized labeled multi-Bernoulli RFSs based on which a novel multi-target tracking filter is proposed and demonstrated via a numerical example. Concluding remarks and extensions are discussed in section VII.

II. Background

This section provides background on conjugate priors and the two key Bayesian inference problems investigated in this paper: The multi-object Bayes update and Chapman-Kolmogorov equation.

A. Conjugate Priors

In Bayesian inference, the hidden *state* (or parameter) $x \in \mathbb{X}$ is assumed to be distributed according to a *prior* p. Further, the state is partially observed as $z \in \mathbb{Z}$, called the *observation*, in accordance with the *likelihood function* $g(z|x)$. All information about the state is encapsulated in the *posterior*, given by Bayes rule:

$$p(x|z) = \frac{g(z|x)p(x)}{\int g(z|\zeta)p(\zeta)d\zeta}.$$

For a given likelihood function, if the posterior $p(\cdot|z)$ belongs to the same family as the prior p, the prior and posterior are said to be *conjugate* distributions, and the prior is called a *conjugate prior*. For example, the Gaussian family is conjugate with respect to a Gaussian likelihood function. Other well-known likelihood-prior combinations in the exponential family include, binomial-beta, Poisson-gamma, and gamma-gamma models, see [15] for a catalogue.

Computing the posterior is generally intractable due to the integration in the normalizing constant $\int g(z|\zeta)p(\zeta)d\zeta$. This is especially true in nonparametric inference where the posterior can be very complex. A conjugate prior is an algebraic convenience, providing a closed-form for the posterior and avoids a difficult numerical integration problem. Moreover, posteriors with the same functional form as the prior often inherit desirable properties that are important for analysis and interpretation.

B. Bayesian Multi-object Inference

This work considers a more general setting where, instead of a single state $x \in \mathbb{X}$, we have a finite set of states $X \subset \mathbb{X}$, called a *multi-object state*, distributed according to a *multi-object prior* π. The multi-object state is partially observed as a finite set of points $Z \subset \mathbb{Z}$, called the *multi-object observation*, through thinning of misdetected objects, Markov shifts of detected objects, and superposition of false observations. The observation model is described by the *multi-object likelihood function* $g(Z|X)$, (given in subsection IV-C) which encapsulates, in addition to the usual observation noise,

- detection uncertainty: each object may or may not generate an observation,
- clutter: observations are corrupted by spurious/false measurements not originating from any object,
- data association uncertainty: there is no information on which object generated which observation.

In this setting, the number of elements of a multi-object state and the values of these elements are random variables. All information on the multi-object state (including the number of objects) is contained in the *multi-object posterior*, given by

$$\pi(X|Y) = \frac{g(Y|X)\pi(X)}{\int g(Y|X)\pi(X)\delta X} \tag{1}$$

where

$$\int f(X)\delta X = \sum_{i=0}^{\infty} \frac{1}{i!} \int_{\mathbb{X}^i} f(\{x_1, ..., x_i\})d(x_1, \cdots, x_i),$$

is the *set integral* of a function f taking the class of finite subsets of $\mathbb{X}$, denoted as $\mathcal{F}(\mathbb{X})$, to the real line $\mathbb{R}$. It is implicit that the multi-object state and observation are modeled as *random finite sets* (RFSs) and that the probability densities of these RFS are interpreted via the Finite Set Statistic (FISST) notion of integration/density proposed by Mahler [1], [20].

While the multi-object posterior is central to Bayesian inference, computationally tractable analytic characterizations are not available in general[1]. In spatial statistics, the multi-object posterior encapsulates the spatial distribution of point patterns as well as relevant statistics [5], [6], and is often approximated via Markov Chain Monte Carlo simulation [7]. In target tracking, the multi-object posterior contains information on target number, locations, velocities [1], and is often approximated, e.g. Probability Hypothesis Density (PHD) filter [20], Cardinalized PHD filter [21], and multi-Bernoulli filter [1], [22]. Implementations of these filters are developed in [23], [24], [25], [26], [22], with convergence analysis given in [23], [27], [28], [29], [30]. Conjugate priors are important in multi-object inference since the posteriors are extremely complex due to the high dimensionality and combinatorial nature of the problem.

[1] For the special case of separable multi-object likelihood functions, the multi-object posterior can be computed analytically for certain classes of (conjugate) priors [17].

C. Chapman-Kolmogorov Equation

An additional problem arises when the multi-object state evolves in time due to motion, births and deaths of objects, as is the case in multi-target tracking. A given multi-object state X evolves to the multi-object state X_+ at the next time step via thinning of dying objects, Markov shift of surviving objects and superposition of new objects. This multi-object dynamic model is described by the *multi-object (Markov) transition kernel* $f(X_+|X)$, (the specific expression is given in subsection IV-C). Before Bayes rule can be applied to compute the posterior from the next observation, the multi-object prior needs to be predicted forward to account for multi-object evolution via the Chapman-Kolmogorov equation [1]

$$\pi_+(X_+) = \int f(X_+|X)\pi(X)\delta X. \tag{2}$$

Bayesian inference in a dynamic setting involves computation of the posterior at each time step. In applications such as multi-target tracking where inference is performed online, recursive computation (of the posterior) via the so-called *multi-object Bayes filter* is the preferred approach. Let X_k and Z_k denote the multi-object state and measurement at time k. Then introducing time indices into the Chapman-Kolmogorov equation (2) and Bayes rule (1), we have, respectively, the prediction and update steps of the multi-object Bayes filter:

$$\pi_{k|k-1}(X_k|Z_{1:k-1}) = \int f_{k|k-1}(X_k|X)\pi_{k-1}(X|Z_{1:k-1})\delta X, \tag{3}$$

$$\pi_k(X_k|Z_{1:k}) = \frac{g_k(Z_k|X_k)\pi_{k|k-1}(X_k|Z_{1:k-1})}{\int g_k(Z_k|X)\pi_{k|k-1}(X|Z_{1:k-1})\delta X}, \tag{4}$$

where $Z_{1:k} = (Z_1, ..., Z_k)$ denotes the measurement history up to time k; $\pi_{k|k-1}(\cdot|Z_{1:k-1})$ and $f_{k|k-1}(\cdot|\cdot)$ denote multi-object prediction and transition kernel from time $k-1$ to k; $\pi_k(\cdot|Z_{1:k})$ and $g_k(\cdot|\cdot)$ denote the multi-object posterior and likelihood function at time k.

In general, the multi-object prediction π_+ is difficult to characterize analytically and does not necessarily take on the same form as the multi-object prior. A known exception is the multi-Bernoulli prior, which preserves its form when predicted forward by a special case of the multi-object transition kernel with multi-Bernoulli births [1]. However, it is not a conjugate prior and the posterior is no-longer multi-Bernoulli. A conjugate multi-object prior family that is closed under the Chapman-Kolmogorov equation means that at all time, the posterior stays in the same family.

III. Labeled Random Finite Sets

This section introduces the notion of labeled RFSs. We begin in subsection III-A with examples of common RFSs and proceed to the treatment of labeled RFSs in subsection III-B.

A. Random finite sets

In essence, an RFS is simply a finite-set-valued random variable. What distinguishes an RFS from a random vector is that: the number of points is random; the points themselves are random and unordered. As mentioned earlier, we use the FISST notion of integration/density to characterize RFSs [1], [20]. While not a probability density [1], the FISST density is equivalent to a probability density relative to an unnormalized distribution of a Poisson RFS (see [23]). For simplicity, in this paper, we shall not distinguish a FISST density and a probability density.

Throughout the paper, we use the standard inner product notation

$$\langle f, g \rangle \triangleq \int f(x)g(x)dx,$$

and the following multi-object exponential notation

$$h^X \triangleq \prod_{x\in X} h(x), \tag{5}$$

where h is a real-valued function, with $h^{\emptyset} = 1$ by convention. We denote a generalization of the Kroneker delta that takes arbitrary arguments such as sets, vectors, integers etc, by

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

and the inclusion function, a generalization of the indicator function, by

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

We also write $1_Y(x)$ in place of $1_Y(\{x\})$ when $X = \{x\}$.

1) Poisson RFS: An RFS X on $\mathbb{X}$ is said to be *Poisson* with *intensity function* v (defined on $\mathbb{X}$) if

- its cardinality $|X|$ is Poisson distributed with mean $\langle v, 1\rangle$, i.e. $|X| \sim \text{Pois}_{\langle v,1\rangle}$, and
- for any finite cardinality, the elements of X are independently and identically distributed (i.i.d.) according to the probability density $v(\cdot)/\langle v, 1\rangle$.

The probability density of this Poisson RFS is given by (see [1] pp. 366)

$$\pi(X) = e^{-\langle v,1\rangle} v^X \tag{6}$$

The intensity function, also known in the tracking literature as the *Probability Hypothesis Density* (PHD), completely characterizes a Poisson RFS.

2) Bernoulli RFS: A *Bernoulli* RFS X on $\mathbb{X}$ has probability $1 - r$ of being empty, and probability r of being a singleton whose (only) element is distributed according to a probability density p (defined on $\mathbb{X}$). The cardinality distribution of a Bernoulli RFS is a Bernoulli distribution with parameter r. The probability density of a Bernoulli RFS is given by (see [1] pp. 368)

$$\pi(X) = \begin{cases} 1 - r & X = \emptyset, \\ r \cdot p(x) & X = \{x\}. \end{cases} \tag{7}$$

3) Multi-Bernoulli RFS: A *multi-Bernoulli* RFS X on $\mathbb{X}$ is a union of a fixed number of independent Bernoulli RFSs $X^{(i)}$ with existence probability $r^{(i)} \in (0, 1)$ and probability density $p^{(i)}$ (defined on $\mathbb{X}$), $i = 1, ..., M$, i.e. $X = \bigcup_{i=1}^{M} X^{(i)}$. A multi-Bernoulli RFS is thus completely described by the multi-Bernoulli parameter set $\{(r^{(i)}, p^{(i)})\}_{i=1}^{M}$. Moreover, its probability density is given by (see [1] pp. 368)

$$\pi(\{x_1, ..., x_n\}) = \prod_{j=1}^{M}(1 - r^{(j)}) \sum_{1\leq i_1 \neq ... \neq i_n \leq M} \prod_{j=1}^{n} \frac{r^{(i_j)} p^{(i_j)}(x_j)}{1 - r^{(i_j)}}. \quad (8)$$

Through out this paper, probability density of the form (8) is abbreviated by $\pi = \{(r^{(i)}, p^{(i)})\}_{i=1}^{M}$.

B. Labeled RFS

In applications involving object identity, for example estimation of (multiple) tracks or trajectories, objects are uniquely identified by a (unobserved) label or tag drawn from a discrete countable space $\mathbb{L} = \{\alpha_i : i \in \mathbb{N}\}$, where $\mathbb{N}$ denotes the set of positive integers and the α_i's are distinct. To incorporate object identity, we augment a mark $\ell \in \mathbb{L}$ to the state $x \in \mathbb{X}$ of each object and consider the multi-object state as a finite set on $\mathbb{X} \times \mathbb{L}$. However, this idea alone is not enough since $\mathbb{L}$ is discrete and it is possible (with non-zero probability) that many targets have the same identity. This problem can be alleviated by the introduction of *labeled RFSs*, which are in essence marked RFSs with distinct marks.

We adhere to the convention that single-object states are represented by lowercase letters, e.g. x, $\mathbf{x}$ while multi-object states are represented by uppercase letters, e.g. X, $\mathbf{X}$. Symbols for labeled states and their distributions/statistics (single-object or multi-object) are bolded to distinguish them from unlabeled ones, e.g. $\mathbf{x}$, $\mathbf{X}$, $\boldsymbol{\pi}$, etc. Spaces are represented by blackboard bold e.g. $\mathbb{X}$, $\mathbb{Z}$, $\mathbb{L}$, $\mathbb{N}$, etc.

Definition 1 *Let $\mathcal{L} : \mathbb{X}\times\mathbb{L} \to \mathbb{L}$ be the projection $\mathcal{L}((x, \ell)) = \ell$, and hence $\mathcal{L}(\mathbf{X}) = \{\mathcal{L}(\mathbf{x}) : \mathbf{x} \in \mathbf{X}\}$ is the set of labels of* $\mathbf{X}$. *A* labeled *RFS with state space $\mathbb{X}$ and (discrete) label space $\mathbb{L}$ is an RFS on $\mathbb{X}\times\mathbb{L}$ such that each realization $\mathbf{X}$ has distinct labels, i.e.*

$$|\mathcal{L}(\mathbf{X})| = |\mathbf{X}|$$

The unlabeled version of a labeled RFS distributed according to $\boldsymbol{\pi}$, is distributed according the marginal

$$\pi(\{x_1, ..., x_n\}) = \sum_{(\ell_1,...,\ell_n)\in\mathbb{L}^n} \boldsymbol{\pi}(\{(x_1, \ell_1), ..., (x_n, \ell_n)\}) \quad (9)$$

The unlabeled version of a labeled RFS is its projection from $\mathbb{X} \times \mathbb{L}$ into $\mathbb{X}$, and is obtained by simply discarding the labels. Intuitively, the cardinality distribution (the distribution of the number of objects) of a labeled RFS should be the same as its unlabeled version. This is formalized by the following Proposition, the proof is straightforward via marginalization, and making use of (9).

Proposition 2 *A labeled RFS and its unlabeled version have the same cardinality distribution.*

For any state space $\mathbb{X}$, let $\mathcal{F}(\mathbb{X})$ denote the collection of finite subsets of $\mathbb{X}$, and $\mathcal{F}_n(\mathbb{X})$ denote the collection of finite subsets of $\mathbb{X}$ with exactly n elements. Since the label space $\mathbb{L}$ is discrete, the set integral for a function $f : \mathcal{F}(\mathbb{X} \times \mathbb{L}) \to \mathbb{R}$ is given by

$$\int f(\mathbf{X})\delta\mathbf{X} = \sum_{i=0}^{\infty} \frac{1}{i!} \sum_{(\ell_1,...,\ell_i)\in\mathbb{L}^i} \int_{\mathbb{X}^i} f(\{(x_1, \ell_1), ..., (x_i, \ell_i)\}) d(x_1, ..., x_i).$$

Throughout the paper it is implicitly assumed that such set integrals exist and are finite. The following lemma is useful for evaluating set integrals involving labeled RFSs (the proof is given in the Appendix.).

Lemma 3 *Let $\Delta(\mathbf{X})$ denote the distinct label indicator $\delta_{|\mathbf{X}|}(|\mathcal{L}(\mathbf{X})|)$. Then, for $h : \mathcal{F}(\mathbb{L}) \to \mathbb{R}$ and $g : \mathbb{X} \times \mathbb{L} \to \mathbb{R}$, integrable on $\mathbb{X}$,*

$$\int \Delta(\mathbf{X})h(\mathcal{L}(\mathbf{X}))g^{\mathbf{X}}\delta\mathbf{X} = \sum_{L\subseteq\mathbb{L}} h(L) \left[\int g(x, \cdot)dx\right]^L$$

Next we present examples of labeled RFS pertinent to the main results of this paper.

1) Labeled Poisson RFS: A labeled Poisson RFS $\mathbf{X}$ with state space $\mathbb{X}$ and label space $\mathbb{L} = \{\alpha_i : i \in \mathbb{N}\}$, is a Poisson RFS X on $\mathbb{X}$ with intensity v on $\mathbb{X}$, tagged or augmented with labels from $\mathbb{L}$. A sample from such labeled Poisson RFS can be generated by the following procedure:

Sampling a labeled Poisson RFS

- initialize $\mathbf{X} = \emptyset$;
- sample $n \sim \text{Pois}_{\langle v,1\rangle}$;
- for $i = 1 : n$
- sample $x \sim v(\cdot)/\langle v, 1\rangle$;
- set $\mathbf{X} = \mathbf{X} \cup \{(x, \alpha_i)\}$;
- end;

This procedure always generates a finite set of augmented states with distinct labels. Intuitively, the set of unlabeled states generated by the above procedure is a Poisson RFS with intensity v. However, the set of labeled states is not a Poisson RFS on $\mathbb{X}\times\mathbb{L}$, in fact its density is given by

$$\boldsymbol{\pi}(\{(x_1, \ell_1), ..., (x_n, \ell_n)\}) = \delta_{\mathbb{L}(n)}(\{\ell_1, ...\ell_n\})\text{Pois}_{\langle v,1\rangle}(n) \prod_{i=1}^{n} \frac{v(x_i)}{\langle v,1\rangle}. \quad (10)$$

where $\text{Pois}_\lambda(n) = e^{-\lambda}\lambda^n/n!$ is the Poisson distribution with rate λ and $\mathbb{L}(n) = \{\alpha_i \in \mathbb{L}\}_{i=1}^{n}$.

To verify the density (10) of the labeled Poisson RFS, consider the likelihood (probability density) that the above procedure generates the points $(x_1, \ell_1), ..., (x_n, \ell_n)$ in that order:

$$\mathbf{p}((x_1, \ell_1), ..., (x_n, \ell_n)) = \delta_{(\alpha_1,...,\alpha_n)}((\ell_1, ...\ell_n))\text{Pois}_{\langle v,1\rangle}(n) \prod_{i=1}^{n} \frac{v(x_i)}{\langle v, 1\rangle}.$$

Following [31], $\boldsymbol{\pi}(\{(x_1,\ell_1),...,(x_n,\ell_n)\})$ is the symmetrization of $\mathbf{p}((x_1,\ell_1),...,(x_n,\ell_n))$ over all permutations σ of $\{1,...,n\}$, i.e.

$$\begin{aligned}&\boldsymbol{\pi}(\{(x_1,\ell_1),...,(x_n,\ell_n)\})\\&=\sum_{\sigma}\mathbf{p}\left((x_{\sigma(1)},\ell_{\sigma(1)}),...,(x_{\sigma(n)},\ell_{\sigma(n)})\right)\\&=\mathrm{Pois}_{\langle v,1\rangle}(n)\sum_{\sigma}\left(\prod_{i=1}^{n}\frac{v(x_{\sigma(i)})}{\langle v,1\rangle}\right)\delta_{(\alpha_1,...,\alpha_n)}((\ell_{\sigma(1)},...,\ell_{\sigma(n)}))\end{aligned}$$

The sum over the permutations is zero, except for the case $\{\ell_1,...\ell_n\}=\{\alpha_1,...,\alpha_n\}$ where there exists a permutation σ such that $(\ell_{\sigma(1)},...\ell_{\sigma(n)})=(\alpha_1,...,\alpha_n)$. Henceforth, (10) follows.

To verify that the unlabeled version of a labeled Poisson RFS is indeed a Poisson RFS, we apply (9) to (10), and simplifying the sum over the labels to give (6).

Remark: The labeled Poisson RFS can be generalized to the labeled *i.i.d. cluster* RFS by removing the Poisson assumption on the cardinality and specifying an arbitrary cardinality distribution.

2) Labeled multi-Bernoulli RFS: A labeled multi-Bernoulli RFS $\mathbf{X}$ with state space $\mathbb{X}$, label space $\mathbb{L}$ and (finite) parameter set $\{(r^{(\zeta)},p^{(\zeta)}):\zeta\in\Psi\}$, is a multi-Bernoulli RFS on $\mathbb{X}$ augmented with labels corresponding to the successful (non-empty) Bernoulli components, i.e. if the Bernoulli component $(r^{(\zeta)},p^{(\zeta)})$ yeilds a non-empty set, then the label of the corresponding state is given by $\alpha(\zeta)$, where $\alpha:\Psi\to\mathbb{L}$ is a 1-1 mapping. The following procedure illustrates how a sample from such a labeled multi-Bernoulli RFS is generated:

Sampling a labeled Multi-Bernoulli RFS

- initialize $\mathbf{X}=\emptyset$;
- for $\zeta\in\Psi$
- sample $u\sim\mathrm{Uniform}[0,1]$;
- if $u\le r^{(\zeta)}$,
- sample $x\sim p^{(\zeta)}(\cdot)$;
- set $\mathbf{X}=\mathbf{X}\cup\{(x,\alpha(\zeta))\}$;
- end;
- end;

It is clear that the above procedure always generates a finite set of augmented states with distinct labels. Intuitively, the set of unlabeled states is a multi-Bernoulli RFS. However, the set of labeled states is not a multi-Bernoulli RFS on $\mathbb{X}\times\mathbb{L}$, in fact its density is given by

$$\begin{aligned}\boldsymbol{\pi}(\{(x_1,\ell_1),...,(x_n,\ell_n)\})=\delta_n(|\{\ell_1,...,\ell_n\}|)\prod_{\zeta\in\Psi}(1-r^{(\zeta)})\\\times\prod_{j=1}^{n}\frac{1_{\alpha(\Psi)}(\ell_j)r^{(\alpha^{-1}(\ell_j))}p^{(\alpha^{-1}(\ell_j))}(x_j)}{1-r^{(\alpha^{-1}(\ell_j))}}.\end{aligned}\tag{11}$$

To verify the density of the labeled multi-Bernoulli RFS (11), consider the likelihood (probability density) that the above procedure generates the points $(x_1,\ell_1),...,(x_n,\ell_n)$ in that order:

$$\begin{aligned}\mathbf{p}((x_1,\ell_1),...,(x_n,\ell_n))&=\delta_n(|\{\ell_1,...\ell_n\}|)\mathrm{ord}(\ell_1,...,\ell_n)\\&\times\prod_{\zeta\in\Psi}(1-r^{(\zeta)})\prod_{j=1}^{n}\frac{1_{\alpha(\Psi)}(\ell_j)r^{(\alpha^{-1}(\ell_j))}p^{(\alpha^{-1}(\ell_j))}(x_j)}{1-r^{(\alpha^{-1}(\ell_j))}}\end{aligned}$$

where $\mathrm{ord}(\ell_1,...,\ell_n)=1$ if $\ell_1<...<\ell_n$ and zero otherwise (since $\mathbb{L}$ is discrete, it is always possible to define an ordering of its elements). Following [31], $\boldsymbol{\pi}(\{(x_1,\ell_1),...,(x_n,\ell_n)\})$ is the symmetrization of $\mathbf{p}((x_1,\ell_1),...,(x_n,\ell_n))$ over all permutations σ of $\{1,...,n\}$, i.e.

$$\begin{aligned}&\boldsymbol{\pi}(\{(x_1,\ell_1),...,(x_n,\ell_n)\})=\\&\sum_{\sigma}\delta_n(|\{\ell_{\sigma(1)},...,\ell_{\sigma(n)}\}|)\mathrm{ord}(\ell_{\sigma(1)},...,\ell_{\sigma(n)})\prod_{\zeta\in\Psi}(1-r^{(\zeta)})\\&\times\prod_{j=1}^{n}\frac{1_{\alpha(\Psi)}(\ell_{\sigma(j)})r^{(\alpha^{-1}(\ell_{\sigma(j)}))}p^{(\alpha^{-1}(\ell_{\sigma(j)}))}(x_{\sigma(j)})}{1-r^{(\alpha^{-1}(\ell_{\sigma(j)}))}}.\end{aligned}$$

If the labels are distinct, then there is only one permutation with the correct ordering and the sum over σ reduces to one term. Moreover, since $\delta_n(|\{\ell_1,...,\ell_n\}|)$ and the product over n is symmetric (permutation invariant) in $(x_1,\ell_1),...,(x_n,\ell_n)$ we have (11).

By combining the two products into one and canceling appropriate terms, the following alternative form of the labeled multi-Bernoulli density can be obtained

$$\boldsymbol{\pi}(\mathbf{X})=\Delta(\mathbf{X})1_{\alpha(\Psi)}(\mathcal{L}(\mathbf{X}))\left[\Phi(\mathbf{X};\cdot)\right]^{\Psi}\tag{12}$$

where $\Delta(\mathbf{X})$ is the distinct label indicator (defined in Lemma **3**), $\mathcal{L}(\mathbf{X})$ is the labels of $\mathbf{X}$ (defined in Definition 1) and

$$\begin{aligned}\Phi(\mathbf{X};\zeta)&=\begin{cases}1-r^{(\zeta)}, & \text{if }\alpha(\zeta)\notin\mathcal{L}(\mathbf{X})\\ r^{(\zeta)}p^{(\zeta)}(x), & \text{if }(x,\alpha(\zeta))\in\mathbf{X}\end{cases}\\&=\sum_{(x,\ell)\in\mathbf{X}}\delta_{\alpha(\zeta)}(\ell)r^{(\zeta)}p^{(\zeta)}(x)+\left(1-1_{\mathcal{L}(\mathbf{X})}(\alpha(\zeta))\right)(1-r^{(\zeta)})\end{aligned}$$

Remark: The equivalence of the two expressions for $\Phi(\mathbf{X};\zeta)$ can be verified by noting that the sum above has only one non-zero term that either takes on $1-r^{(\zeta)}$ if $\alpha(\zeta)\notin\mathcal{L}(\mathbf{X})$, or $r^{(\zeta)}p^{(\zeta)}(x)$ if $(x,\alpha(\zeta))\in\mathbf{X}$. The second expression for $\Phi(\mathbf{X};\zeta)$ is useful in the multi-object Chapman-Kolmogorov equation.

To verify that the unlabeled version of a labeled multi-Bernoulli RFS is indeed a multi-Bernoulli RFS, we apply (9) to (11) and simplify the sum over the labels.

IV. GENERALIZED LABELED MULTI-BERNOULLI

In this section, we introduce a family of labeled RFS distributions which is conjugate with respect to the multi-object likelihood function, and show that it is closed under the multi-object Chapman-Kolmogorov equation with respect to the multi-object transition kernel.

A. Generalized labeled Multi-Bernoulli RFS

Definition 4 *A generalized labeled multi-Bernoulli RFS is a labeled RFS with state space $\mathbb{X}$ and (discrete) label space $\mathbb{L}$ distributed according to*

$$\boldsymbol{\pi}(\mathbf{X})=\Delta(\mathbf{X})\sum_{c\in\mathbb{C}}w^{(c)}(\mathcal{L}(\mathbf{X}))\left[p^{(c)}\right]^{\mathbf{X}} \quad (13)$$

where $\mathbb{C}$ is a discrete index set, $w^{(c)}(L)$ and $p^{(c)}$ satisfy

$$\sum_{L\subseteq\mathbb{L}}\sum_{c\in\mathbb{C}}w^{(c)}(L) = 1, \quad (14)$$

$$\int p^{(c)}(x,\ell)dx = 1. \quad (15)$$

A generalized labeled multi-Bernoulli can be interpreted as a mixture of multi-object exponentials. Each term in the mixture (13) consists of a weight $w^{(c)}(\mathcal{L}(\mathbf{X}))$ that only depends on the labels of the multi-object state, and a multi-object exponential $\left[p^{(c)}\right]^{\mathbf{X}}$ that depends on the entire multi-object state. The points of a generalized labeled multi-Bernoulli RFS are not statistically independent.

The following results shows the cardinality distribution and PHD (or intensity function) of generalized labeled multi-Bernoulli RFS. Proofs are given in the Appendix.

Proposition 5 *The cardinality distribution of a generalized labeled multi-Bernoulli RFS is given by*

$$\boldsymbol{\rho}(n)=\sum_{L\in\mathcal{F}_n(\mathbb{L})}\sum_{c\in\mathbb{C}}w^{(c)}(L) \quad (16)$$

From property (14), it can be easily verified that (16) is indeed a probability distribution,

$$\begin{aligned}\sum_{n=0}^{\infty}\boldsymbol{\rho}(n) &= \sum_{n=0}^{\infty}\sum_{L\in\mathcal{F}_n(\mathbb{L})}\sum_{c\in\mathbb{C}}w^{(c)}(L)\\ &= \sum_{L\subseteq\mathbb{L}}\sum_{c\in\mathbb{C}}w^{(c)}(L)=1,\end{aligned}$$

and that the probability density $\boldsymbol{\pi}$ in (13) integrates to 1,

$$\begin{aligned}\int\boldsymbol{\pi}(\mathbf{X})\delta\mathbf{X} &= \sum_{n=0}^{\infty}\frac{1}{n!}\int_{(\mathbb{L}\times\mathbb{X})^n}\boldsymbol{\pi}(\{\mathbf{x}_1,...,\mathbf{x}_n\})d(\mathbf{x}_1...,\mathbf{x}_n)\\ &= \sum_{n=0}^{\infty}\boldsymbol{\rho}(n)\\ &= 1.\end{aligned}$$

Hence, the RFS defined above is indeed a labeled RFS since $\Delta(\mathbf{X})=\delta_{|\mathbf{X}|}(|\mathcal{L}(\mathbf{X})|)$ is the distinct-label indicator of $\mathbf{X}$, i.e. takes on the value 1 if the labels of $\mathbf{X}$ are distinct and zero otherwise.

Proposition 6 *The PHD (or intensity function) of the unlabeled version of a generalized labeled multi-Bernoulli RFS is given by*

$$v(x)=\sum_{c\in\mathbb{C}}\sum_{\ell\in\mathbb{L}}p^{(c)}(x,\ell)\sum_{L\subseteq\mathbb{L}}1_L(\ell)w^{(c)}(L) \quad (17)$$

To verify that the PHD mass is equal to the mean cardinality, note that each $p^{(c)}(\cdot,\ell)$ integrates to 1, $\sum_{L\subseteq\mathbb{L}}$ is equivalent to $\sum_{n=0}^{\infty}\sum_{L\in\mathcal{F}_n(\mathbb{L})}$ and exchanging the order of the summations, we have

$$\begin{aligned}\int v(x)dx &= \sum_{n=0}^{\infty}\sum_{c\in\mathbb{C}}\sum_{L\in\mathcal{F}_n(\mathbb{L})}\left[\sum_{\ell\in\mathbb{L}}1_I(\ell)\right]w^{(c)}(I)\\ &= \sum_{n=0}^{\infty}\sum_{c\in\mathbb{C}}\sum_{L\in\mathcal{F}_n(\mathbb{L})}nw^{(c)}(L)\\ &= \sum_{n=0}^{\infty}n\boldsymbol{\rho}(n)\end{aligned}$$

B. Special cases

Generalized labeled multi-Bernoulli RFSs cover labeled Poisson and labeled multi-Bernoulli RFSs discussed earlier. For the labeled Poisson RFS density (10), note that $\delta_{\mathbb{L}(|\mathbf{X}|)}(\mathcal{L}(\mathbf{X}))=\Delta(\mathbf{X})\delta_{\mathbb{L}(|\mathcal{L}(\mathbf{X})|)}(\mathcal{L}(\mathbf{X}))$, and hence

$$\boldsymbol{\pi}(\mathbf{X})=\Delta(\mathbf{X})\delta_{\mathbb{L}(|\mathcal{L}(\mathbf{X})|)}(\mathcal{L}(\mathbf{X}))\mathrm{Pois}_{\langle v,1\rangle}(|\mathcal{L}(\mathbf{X})|)\prod_{(x,\ell)\in\mathbf{X}}\frac{v(x)}{\langle v,1\rangle}$$

Thus, the labeled Poisson RFS is a special case of the generalized labeled multi-Bernoulli RFS with

$$\begin{aligned}p^{(c)}(x,\ell) &= v(x)/\langle v,1\rangle\\ w^{(c)}(L) &= \delta_{\mathbb{L}(|L|)}(L)\mathrm{Pois}_{\langle v,1\rangle}(|L|).\end{aligned}$$

Note that the index space $\mathbb{C}$ has only one element, in which case the (c) superscript is unnecessary.

The labeled multi-Bernoulli RFS distributed by (11) is a special case of the generalized labeled multi-Bernoulli RFS with

$$\begin{aligned}p^{(c)}(x,\ell) &= p^{(\alpha^{-1}(\ell))}(x)\\ w^{(c)}(L) &= \prod_{\zeta\in\Psi}(1-r^{(\zeta)})\prod_{\ell\in L}\frac{1_{\alpha(\Psi)}(\ell)r^{(\alpha^{-1}(\ell))}}{1-r^{(\alpha^{-1}(\ell))}}.\end{aligned}$$

Note again, that the index space $\mathbb{C}$ has only one element, in which case the (c) superscript is unnecessary.

C. Multi-object Conjugate Prior

In this subsection we establish conjugacy of the generalized labeled multi-Bernoulli family with respect to the multi-object likelihood function that accommodates thinning, Markov shifts, and superposition. A more concise description of the observation model is given as follows.

For a given multi-object state $\mathbf{X}$, each state $\mathbf{x}\in\mathbf{X}$ is either detected with probability $p_D(\mathbf{x})$ and generates a point z with likelihood $g(z|\mathbf{x})$, or missed with probability $q_D(\mathbf{x})=1-p_D(\mathbf{x})$, i.e. $\mathbf{x}$ generates a Bernoulli RFS with parameter $(p_D(\mathbf{x}),g(\cdot|\mathbf{x}))$. Assuming that conditional on $\mathbf{X}$ these Bernoulli RFSs are independent, then the set $W\subset\mathbb{Z}$ of detected points (non-clutter measurements) is a multi-Bernoulli RFS with parameter set $\{(p_D(\mathbf{x}),g(\cdot|\mathbf{x})){:}\mathbf{x}\in\mathbf{X}\}$, i.e. W is distributed according to

$$\pi_D(W|\mathbf{X})=\{(p_D(\mathbf{x}),g(\cdot|\mathbf{x})):\mathbf{x}\in\mathbf{X}\}(W).$$

The set $Y \subset \mathbb{Z}$ of false observations (or clutter), assumed independent of the detected points, is modeled by a Poisson RFS with intensity function $\kappa(\cdot)$, i.e. Y is distributed according to

$$\pi_K(Y) = e^{-\langle \kappa,1\rangle}\kappa^Y.$$

The multi-object observation Z is the superposition of the detected points and false observations, i.e. $Z = W \cup Y$. Hence, it follows from FISST, that the multi-object likelihood is the convolution of π_D and π_K, i.e.

$$g(Z|\mathbf{X}) = \sum_{W\subseteq Z} \pi_D(W|\mathbf{X})\pi_K(Z - W). \tag{18}$$

where $Z - W$ denotes the set difference between Z and W.

Proposition 7 *If the prior distribution is a generalized labeled multi-Bernoulli of the form (13), then, under the multi-object likelihood function (18), the posterior distribution is also a generalized multi-Bernoulli given by*

$$\boldsymbol{\pi}(\mathbf{X}|Z) = \Delta(\mathbf{X})\sum_{c\in\mathbb{C}}\sum_{\theta\in\Theta} w_Z^{(c,\theta)}(\mathcal{L}(\mathbf{X}))\left[p^{(c,\theta)}(\cdot|Z)\right]^{\mathbf{X}} \tag{19}$$

where Θ *is the space of mappings* $\theta : \mathbb{L} \to \{0 : |Z|\} \triangleq \{0, 1, ..., |Z|\}$, *such that* $\theta(i) = \theta(i') > 0$ *implies* $i = i'$,

$$w_Z^{(c,\theta)}(L) = \frac{\delta_{\theta^{-1}(\{0:|Z|\})}(L)w^{(c)}(L)[\eta_Z^{(c,\theta)}]^L}{\sum_{c\in\mathbb{C}}\sum_{\theta\in\Theta}\sum_{J\subseteq\mathbb{L}}\delta_{\theta^{-1}(\{0:|Z|\})}(J)w^{(c)}(J)[\eta_Z^{(c,\theta)}]^J}, \tag{20}$$

$$p^{(c,\theta)}(x,\ell|Z) = \frac{p^{(c)}(x,\ell)\psi_Z(x,\ell;\theta)}{\eta_Z^{(c,\theta)}(\ell)}, \tag{21}$$

$$\eta_Z^{(c,\theta)}(\ell) = \left\langle p^{(c)}(\cdot,\ell), \psi_Z(\cdot,\ell;\theta)\right\rangle, \tag{22}$$

$$\psi_Z(x,\ell;\theta) = \delta_0(\theta(\ell))q_D(x,\ell) + (1-\delta_0(\theta(\ell)))\frac{p_D(x,\ell)g(z_{\theta(\ell)}|x,\ell)}{\kappa(z_{\theta(\ell)})}. \tag{23}$$

Proposition 7 explicitly describes how to calculate the parameters $w_Z^{(c,\theta)}(\cdot)$, $p^{(c,\theta)}(\cdot,\cdot|Z)$ of the updated multi-object density from the parameters $w^{(c)}(\cdot)$, $p^{(c)}(\cdot,\cdot)$ of the prior multi-object density. The domain of the mapping θ is $\theta^{-1}(\{0 : |Z|\})$, i.e. the inverse image of $\{0 : |Z|\}$ under θ, and the term $\delta_{\theta^{-1}(\{0:|Z|\})}(\mathcal{L}(\mathbf{X}))$ in (20) implies that we only need to consider those mappings with domain $\mathcal{L}(\mathbf{X}) \subseteq \mathbb{L}$. The updated single-object density for a given label $p^{(c,\theta)}(\cdot,\ell|Z)$ is computed from the prior single-object density $p^{(c)}(\cdot,\ell)$ via Bayes rule with "likelihood function" $\psi_Z(\cdot,\ell;\theta)$. Note that $\psi_Z(x,\ell;\theta)$ is either $q_D(x,\ell)$ if $\theta(\ell) = 0$, or $p_D(x,\ell)g(z_{\theta(\ell)}|x,\ell)/\kappa(z_{\theta(\ell)})$ if $\theta(\ell) > 0$, i.e.

$$\psi_Z(x,\ell;\theta) = \begin{cases} \frac{p_D(x,\ell)g(z_{\theta(\ell)}|x,\ell)}{\kappa(z_{\theta(\ell)})} & \theta(\ell) > 0 \\ q_D(x,\ell) & \theta(\ell) = 0 \end{cases} \tag{24}$$

For a valid label set L, i.e. $\delta_{\theta^{-1}(\{0:|Z|\})}(L) = 1$, the updated weight $w_Z^{(c,\theta)}(L)$ is proportional to the prior weight $w^{(c)}(L)$ scaled by the product $[\eta_Z^{(c,\theta)}]^L$ of single-object normalizing constants.

D. Multi-object Chapman-Kolmogorov Prediction

In this subsection we show that the generalized labeled multi-Bernoulli family is closed under the Chapman-Kolmogorov prediction for the multi-object transition kernel that involves thinning, Markov shifts and superposition of new targets. A more concise description of the multi-object dynamic model is given as follows.

Each object is identified by a unique label $\ell = (k, i)$, where k is the time of birth and i is a unique index to distinguish objects born at the same time. The label space $\mathbb{L}_k$ for objects born at time k is then $\{k\} \times \mathbb{N}$ and the label space $\mathbb{L}_{0:k}$ for objects at time k (including those born prior to k) can be constructed recursively by $\mathbb{L}_{0:k} = \mathbb{L}_{0:k-1} \cup \mathbb{L}_k$. An object born at time k, has state $\mathbf{x} \in \mathbb{X}\times\mathbb{L}_k$ and a multi-object state $\mathbf{X}$ at time k, is a finite subset of $\mathbb{X}\times\mathbb{L}_{0:k}$. This construct implies that we can distinguish surviving and birth objects from their labels. Without explicit reference to the time index, let $\mathbb{L}$ denote the label space for the current time, $\mathbb{B}$ the label space for objects born at the next time, and $\mathbb{L}_+ = \mathbb{L} \cup \mathbb{B}$ the label space for the next time. Note that $\mathbb{L}$ and $\mathbb{B}$ are disjoint i.e. $\mathbb{L} \cap \mathbb{B} = \emptyset$.

Given the current multi-object state $\mathbf{X}$, each state $(x, \ell) \in \mathbf{X}$ either continues to exist at the next time step with probability $p_S(x, \ell)$ and moves to a new state (x_+, ℓ_+) with probability density $f(x_+|x, \ell)\delta_\ell(\ell_+)$, or dies with probability $q_S(x, \ell) = 1 - p_S(x, \ell)$. Note that the label or identity of the object is preserved in the transition, only the kinematic (non-label) part of state changes. Assuming that $\mathbf{X}$ has distinct labels and that conditional on $\mathbf{X}$, the transition of the kinematic states are mutually independent, then the set $\mathbf{W}$ of surviving objects at the next time is a labeled multi-Bernoulli RFS with parameter set $\{(p_S(\mathbf{x}), f(\cdot|\mathbf{x})) : \mathbf{x} \in \mathbf{X}\}$ and labeling function $\alpha : \mathbf{X} \to \mathbb{L}$ defined by $\alpha(x, \ell) = \ell$. Hence, it follows from (12) that $\mathbf{W}$ is distributed according to

$$\mathbf{f}_S(\mathbf{W}|\mathbf{X}) = \Delta(\mathbf{W})\Delta(\mathbf{X})1_{\mathcal{L}(\mathbf{X})}(\mathcal{L}(\mathbf{W}))\left[\Phi(\mathbf{W};\cdot)\right]^{\mathbf{X}} \tag{25}$$

where

$$\Phi(\mathbf{W}; x,\ell) = \sum_{(x_+,\ell_+)\in\mathbf{W}} \delta_\ell(\ell_+)p_S(x,\ell)f(x_+|x,\ell) + [1-1_{\mathcal{L}(\mathbf{W})}(\ell)]q_S(x,\ell)$$

The $\Delta(\mathbf{X})$ in (25) ensures that only $\mathbf{X}$ with distinct labels are considered.

The set $\mathbf{Y}$ of new born objects at the next time is modeled as a labeled RFS with label space $\mathbb{B}$, distributed according to $\mathbf{f}_B$. Note that $\mathbf{f}_B(\mathbf{Y}) = 0$ if $\mathbf{Y}$ contains an element $\mathbf{y}$ with $\mathcal{L}(\mathbf{y}) \notin \mathbb{B}$. Examples of labeled births RFS are labeled Poisson, labeled multi-Bernoulli or generalized label multi-Bernoulli. Without loss of generality we use the following birth density through out:

$$\mathbf{f}_B(\mathbf{Y}) = \Delta(\mathbf{Y})w_B(\mathcal{L}(\mathbf{Y}))\left[p_B\right]^{\mathbf{Y}} \tag{26}$$

Our result readily extends to generalized labeled multi-Bernoulli birth, albeit the prediction density $\boldsymbol{\pi}_+$ becomes rather cumbersome. The birth model (26) covers both labeled Poisson with

$$w_B(J) = \delta_{\mathbb{B}(|J|)}(J)\text{Pois}_{\langle\gamma,1\rangle}(|J|), \quad (27)$$
$$p_B(x,\ell) = \gamma(x)/\langle\gamma,1\rangle, \quad (28)$$

and labeled multi-Bernoulli with

$$w_B(J) = \prod_{i\in\mathbb{B}(M)}\left(1-r_B^{(i)}\right)\prod_{\ell\in J}\frac{1_{\mathbb{B}(M)}(\ell)r_B^{(\ell)}}{1-r_B^{(\ell)}}, \quad (29)$$
$$p_B(x,\ell) = p_B^{(\ell)}(x). \quad (30)$$

The multi-object state at the next time $\mathbf{X}_+$ is the superposition of surviving objects and new born objects, i.e. $\mathbf{X}_+ = \mathbf{W}\cup\mathbf{Y}$. Since the label spaces $\mathbb{L}$ and $\mathbb{B}$ are disjoint, the (labeled) birth objects and surviving objects are independent. Hence it follows from FISST that the multi-object transition kernel $\mathbf{f}(\cdot|\cdot) : \mathcal{F}(\mathbb{X}\times\mathbb{L}_+)\times\mathcal{F}(\mathbb{X}\times\mathbb{L}) \to [0,\infty)$ is given by

$$\mathbf{f}(\mathbf{X}_+|\mathbf{X}) = \sum_{\mathbf{W}\subseteq\mathbf{X}_+}\mathbf{f}_S(\mathbf{W}|\mathbf{X})\mathbf{f}_B(\mathbf{X}_+ - \mathbf{W}).$$

Consider the subset of $\mathbf{X}_+$ that consist of surviving objects $\mathbf{X}_+\cap(\mathbb{X}\times\mathbb{L}) = \{\mathbf{x}_+\in\mathbf{X}_+ : \mathcal{L}(\mathbf{x}_+)\in\mathbb{L}\}$. For any $\mathbf{W}\subseteq\mathbf{X}_+$, if $\mathbf{W}$ is not a subset of the surviving objects, i.e. $\mathbf{W}\nsubseteq\mathbf{X}_+\cap\mathbb{X}\times\mathbb{L}$, then the (set of) labels of $\mathbf{W}$ is not a subset of current labels space, i.e. $\mathcal{L}(\mathbf{W})\nsubseteq\mathbb{L}$, and $1_{\mathcal{L}(\mathbf{X})}(\mathcal{L}(\mathbf{W})) = 0$, consequently it follows from (25) that $\mathbf{f}_S(\mathbf{W}|\mathbf{X}) = 0$. Hence, we only need to consider $\mathbf{W}\subseteq\mathbf{X}_+\cap(\mathbb{X}\times\mathbb{L})$. Furthermore, for any non-empty $\mathbf{W}$, if $\mathbf{W}\subset\mathbf{X}_+\cap(\mathbb{X}\times\mathbb{L})$, then there exists $\mathbf{x}_+\in\mathbf{X}_+ - \mathbf{W}$ such that $\mathcal{L}(\mathbf{x}_+)\in\mathbb{L}$, i.e. $\mathcal{L}(\mathbf{x}_+)\notin\mathbb{B}$ (since $\mathbb{L}$ and $\mathbb{B}$ are disjoint) and hence $\mathbf{f}_B(\mathbf{X}_+ - \mathbf{W}) = 0$. Consequently, the transition kernel reduces to the product of the transition density for surviving objects and the density of new objects.

$$\mathbf{f}(\mathbf{X}_+|\mathbf{X}) = \mathbf{f}_S(\mathbf{X}_+\cap(\mathbb{X}\times\mathbb{L})|\mathbf{X})\mathbf{f}_B(\mathbf{X}_+ - \mathbb{X}\times\mathbb{L}) \quad (31)$$

Proposition 8 *If the current multi-object prior is a generalized labeled multi-Bernoulli of the form (13), then the predicted multi-object density is also a generalized labeled multi-Bernoulli given by*

$$\boldsymbol{\pi}_+(\mathbf{X}_+) = \Delta(\mathbf{X}_+)\sum_{c\in\mathbb{C}}w_+^{(c)}(\mathcal{L}(\mathbf{X}_+))\left[p_+^{(c)}\right]^{\mathbf{X}_+} \quad (32)$$

where

$$w_+^{(c)}(L) = w_B(L-\mathbb{L})w_S^{(c)}(L\cap\mathbb{L}), \quad (33)$$
$$p_+^{(c)}(x,\ell) = 1_{\mathbb{L}}(\ell)p_S^{(c)}(x,\ell) + (1-1_{\mathbb{L}}(\ell))p_B(x,\ell), \quad (34)$$
$$p_S^{(c)}(x,\ell) = \frac{\left\langle p_S(\cdot,\ell)f(x|\cdot,\ell),p^{(c)}(\cdot,\ell)\right\rangle}{\eta_S^{(c)}(\ell)}, \quad (35)$$
$$\eta_S^{(c)}(\ell) = \int\left\langle p_S(\cdot,\ell)f(x|\cdot,\ell),p^{(c)}(\cdot,\ell)\right\rangle dx, \quad (36)$$
$$w_S^{(c)}(J) = [\eta_S^{(c)}]^J\sum_{I\subseteq\mathbb{L}}1_I(J)[q_S^{(c)}]^{I-J}w^{(c)}(I), \quad (37)$$
$$q_S^{(c)}(\ell) = \left\langle q_S(\cdot,\ell),p^{(c)}(\cdot,\ell)\right\rangle. \quad (38)$$

Proposition 8 explicitly describes how to calculate the parameters $w_+^{(c)}(\cdot)$, $p_+^{(c)}(\cdot,\cdot)$ of the predicted multi-object density from the parameters $w^{(c)}(\cdot)$, $p^{(c)}(\cdot,\cdot)$ of the prior multi-object density. For a given label set, L, the weight $w_+^{(c)}(L)$ is the product of weight $w_B(L-\mathbb{L})$ of the birth labels $L-\mathbb{L} = L\cap\mathbb{B}$ and the weight $w_S^{(c)}(L\cap\mathbb{L})$ of the surviving labels $L\cap\mathbb{L}$. The weight $w_S^{(c)}(J)$ involves a weighted sum of the prior weights over all label sets that contains the surviving set J. The predicted single-object density for a given label $p_+^{(c)}(\cdot,\ell)$ is either the density $p_B(\cdot,\ell)$ of a newly born object or the density $p_S^{(c)}(\cdot,\ell)$ of a surviving object computed from the prior density $p^{(c)}(\cdot,\ell)$ via the single-object prediction with transition density $f(\cdot|\cdot,\ell)$ weighted by the probability of survival $p_S(\cdot,\ell)$.

In a dynamic multi-object estimation context, Propositions 7, and 8 imply that starting from a generalized labeled multi-Bernoulli initial prior, all subsequent prediction and posterior densities are also generalized labeled multi-Bernoullis. Proposition 8 applies to the prediction step (3) and Proposition 7 applies to the update step (4) of the multi-object Bayes filter.

V. Stronger results

In this section we establish stronger results that are immediately applicable to multi-target tracking. Specifically, in subsection V-A we introduce a smaller family within the class of generalized labeled multi-Bernoulli RFSs that is also closed under the Chapman-Kolmogorov equation and Bayes rule. These results further reduce the computational cost and memory requirements of the prediction and update steps. Subsection VI details a multi-target tracker that can be implemented via Gaussian mixture or sequential Monte Carlo.

A. Stronger results

Definition 9 *A δ-generalized labeled multi-Bernoulli RFS with state space $\mathbb{X}$ and (discrete) label space $\mathbb{L}$ is a special case of a generalized labeled multi-Bernoulli with*

$$\mathbb{C} = \mathcal{F}(\mathbb{L})\times\Xi$$
$$w^{(c)}(L) = w^{(I,\xi)}(L) = \omega^{(I,\xi)}\delta_I(L)$$
$$p^{(c)} = p^{(I,\xi)} = p^{(\xi)}$$

where Ξ is a discrete space, i.e. it is distributed according to

$$\boldsymbol{\pi}(\mathbf{X}) = \Delta(\mathbf{X})\sum_{(I,\xi)\in\mathcal{F}(\mathbb{L})\times\Xi}\omega^{(I,\xi)}\delta_I(\mathcal{L}(\mathbf{X}))\left[p^{(\xi)}\right]^{\mathbf{X}} \quad (39)$$

Remark: In target tracking, a δ-generalized labeled multi-Bernoulli can be used to represent the multi-object prediction density to time k or multi-object filtering density at time k. Each $I\in\mathcal{F}(\mathbb{L})$ represents a set of tracks labels at time k. For the prediction density each $\xi\in\Xi$ represents a history of association maps upto time $k-1$, i.e. $\xi = (\theta_1,...,\theta_{k-1})$, where an association map at time j is a function θ_j which maps track labels at time j to measurement indices at time j. For the filtering density each ξ represents a history of association maps upto time k, i.e. $\xi = (\theta_1,...,\theta_k)$. The pair (I,ξ) represents the *hypothesis* that the set of tracks I has a history ξ of association maps. The weight $\omega^{(I,\xi)}$ represents the probability of hypothesis (I,ξ) and $p^{(\xi)}(\cdot,\ell)$ is the probability density of the kinematic state of track ℓ for the association map history ξ. Note that not all hypotheses are feasible, i.e. not all pairs (I,ξ) are consistent (with each other), and infeasible ones have zero weights.

Observe that for a generalized labeled multi-Bernoulli RFS with $\mathbb{C}=\mathcal{F}(\mathbb{L})\times\Xi$, the number of $w^{(c)}$ and $p^{(c)}$ we need to store/compute are $|\mathcal{F}(\mathbb{L})\times\Xi|$ and $|\mathcal{F}(\mathbb{L})\times\Xi|$, whereas for a δ-generalized labeled multi-Bernoulli RFS the number of $w^{(c)}$ and $p^{(c)}$ we need to store/compute are $|\mathcal{F}(\mathbb{L})\times\Xi|$ and $|\Xi|$. In practice, we can further reduce the computation/storage by approximating $\mathcal{F}(\mathbb{L})\times\Xi$ and Ξ with smaller subsets of feasible hypotheses that have significant weights.

Using Proposition 5, the cardinality distribution of a δ-generalized labeled multi-Bernoulli RFS becomes

$$\begin{aligned}\boldsymbol{\rho}(n) &= \sum_{(I,\xi)\in\mathcal{F}(\mathbb{L})\times\Xi}\ \sum_{L\in\mathcal{F}_n(\mathbb{L})}\omega^{(I,\xi)}\delta_I(L)\\ &= \sum_{(I,\xi)\in\mathcal{F}_n(\mathbb{L})\times\Xi}\omega^{(I,\xi)} \qquad (40)\end{aligned}$$

Thus, the probability of n tracks is the sum of the weights of hypotheses with exactly n tracks. The PHD of the unlabeled version of a δ-generalized labeled multi-Bernoulli RFS reduces to the following form via Proposition 6

$$\begin{aligned}v(x) &= \sum_{(I,\xi)\in\mathcal{F}(\mathbb{L})\times\Xi}\ \sum_{\ell\in\mathbb{L}}p^{(\xi)}(x,\ell)\sum_{L\subseteq\mathbb{L}}1_L(\ell)\omega^{(I,\xi)}\delta_I(L)\\ &= \sum_{\ell\in\mathbb{L}}\ \sum_{(I,\xi)\in\mathcal{F}(\mathbb{L})\times\Xi}\omega^{(I,\xi)}1_I(\ell)p^{(\xi)}(x,\ell)\end{aligned}$$

The inner sum, i.e. the weighted sum of densities for track ℓ over all hypotheses that contain track ℓ, can be interpreted as the PHD of track ℓ, and the sum of the weights $\sum_{(I,\xi)\in\mathcal{F}(\mathbb{L})\times\Xi}\omega^{(I,\xi)}1_I(\ell)$ can be interpreted as the existence probability for track ℓ. The total PHD, is then the sum of the PHDs of all tracks, which can be further simplified to

$$v(x) = \sum_{(I,\xi)\in\mathcal{F}(\mathbb{L})\times\Xi}\omega^{(I,\xi)}\sum_{\ell\in I}p^{(\xi)}(x,\ell) \qquad (41)$$

The following results show that the family of δ-generalized labeled multi-Bernoullis is closed under the Chapman-Kolmogorov prediction and Bayes update (the proofs are relegated to the Appendix).

Proposition 10 *If the multi-object prior is a δ-generalized labeled multi-Bernoulli of the form (39), then the multi-object prediction is also a δ-generalized labeled multi-Bernoulli with the following form*

$$\boldsymbol{\pi}_+(\mathbf{X}_+) = \Delta(\mathbf{X}_+)\sum_{(I_+,\xi)\in\mathcal{F}(\mathbb{L}_+)\times\Xi}\omega_+^{(I_+,\xi)}\delta_{I_+}(\mathcal{L}(\mathbf{X}_+))\left[p_+^{(\xi)}\right]^{\mathbf{X}_+} \qquad (42)$$

where

$$\begin{aligned}\omega_+^{(I_+,\xi)} &= w_B(I_+\cap\mathbb{B})\omega_S^{(\xi)}(I_+\cap\mathbb{L}) &(43)\\ p_+^{(\xi)}(x,\ell) &= 1_{\mathbb{L}}(\ell)p_S^{(\xi)}(x,\ell) + (1-1_{\mathbb{L}}(\ell))p_B(x,\ell) &(44)\\ p_S^{(\xi)}(x,\ell) &= \frac{\langle p_S(\cdot,\ell)f(x|\cdot,\ell),p^{(\xi)}(\cdot,\ell)\rangle}{\eta_S^{(\xi)}(\ell)} &(45)\\ \eta_S^{(\xi)}(\ell) &= \int\left\langle p_S(\cdot,\ell)f(x|\cdot,\ell),p^{(\xi)}(\cdot,\ell)\right\rangle dx &(46)\\ \omega_S^{(\xi)}(L) &= [\eta_S^{(\xi)}]^L\sum_{I\subseteq\mathbb{L}}1_I(L)[q_S^{(\xi)}]^{I-L}\omega^{(I,\xi)} &(47)\\ q_S^{(\xi)}(\ell) &= \left\langle q_S(\cdot,\ell),p^{(\xi)}(\cdot,\ell)\right\rangle &(48)\end{aligned}$$

Remark: Proposition 8 only implies that the predicted generalized labeled multi-Bernoulli is a sum over $(I,\xi)\in\mathcal{F}(\mathbb{L})\times\Xi$, but Proposition 10 is a stronger result asserting that the sum is over $(I_+,\xi)\in\mathcal{F}(\mathbb{L}_+)\times\Xi$ where $\mathbb{L}_+=\mathbb{L}\cup\mathbb{B}$, the space of labels at the next time. The predicted multi-object density $\boldsymbol{\pi}_+$ involves a new double sum over the index $(I_+,\xi)\in\mathcal{F}(\mathbb{L}_+)\times\Xi$ which includes a sum over surviving label sets $I\subset\mathbb{L}$. This is more intuitive from a multi-target tracking point of view as it shows how labels of new objects are introduced into the prediction.

Proposition 11 *If the multi-object prior is a δ-generalized labeled multi-Bernoulli of the form (39), then the multi-object posterior is also a δ-generalized labeled multi-Bernoulli with the following form*

$$\boldsymbol{\pi}(\mathbf{X}|Z) = \Delta(\mathbf{X})\sum_{(I,\xi)\in\mathcal{F}(\mathbb{L})\times\Xi}\ \sum_{\theta\in\Theta}\omega^{(I,\xi,\theta)}(Z)\delta_I(\mathcal{L}(\mathbf{X}))\left[p^{(\xi,\theta)}(\cdot|Z)\right]^{\mathbf{X}} \qquad (49)$$

where Θ is the space of mappings $\theta:\mathbb{L}\to\{0,1,...,|Z|\}$, such that $\theta(i)=\theta(i')>0$ implies $i=i'$, and

$$\omega^{(I,\xi,\theta)}(Z) = \frac{\delta_{\theta^{-1}(\{0:|Z|\})}(I)\omega^{(I,\xi)}[\eta_Z^{(\xi,\theta)}]^I}{\sum\limits_{(I,\xi)\in\mathcal{F}(\mathbb{L})\times\Xi}\sum\limits_{\theta\in\Theta}\delta_{\theta^{-1}(\{0:|Z|\})}(I)\omega^{(I,\xi)}[\eta_Z^{(\xi,\theta)}]^I}, \qquad (50)$$

$$\begin{aligned}p^{(\xi,\theta)}(x,\ell|Z) &= \frac{p^{(\xi)}(x,\ell)\psi_Z(x,\ell;\theta)}{\eta_Z^{(\xi,\theta)}(\ell)}, &(51)\\ \eta_Z^{(\xi,\theta)}(\ell) &= \left\langle p^{(\xi)}(\cdot,\ell),\psi_Z(\cdot,\ell;\theta)\right\rangle, &(52)\\ \psi_Z(x,\ell;\theta) &= \delta_0(\theta(\ell))q_D(x,\ell)\\ &\quad+(1-\delta_0(\theta(\ell)))\frac{p_D(x,\ell)g(z_{\theta(\ell)}|x,\ell)}{\kappa(z_{\theta(\ell)})}. &(53)\end{aligned}$$

VI. Application to Multi-Target Tracking

We now illustrate an application of the established results to multiple target tracking where the prediction (3) and update (4) are recursively applied to propagate the multi-object posterior density forward in time. The proposed filter generalizes the multi-Bernoulli filter in [22] since it is based on generalized labeled multi-Bernoulli RFSs.

Propositions 10 and 11 imply that starting from a δ-generalized labeled multi-Bernoulli initial prior with $\mathbb{L}=\mathbb{L}_0$, $\Xi=\emptyset$, the predicted and updated multi-object densities, at each subsequent time step $k>0$, are δ-generalized labeled multi-Bernoullis with $\mathbb{L}=\mathbb{L}_{0:k}$, $\Xi=\Theta_1\times\cdots\times\Theta_{k-1}$, and $\mathbb{L}=\mathbb{L}_{0:k}$, $\Xi=\Theta_1\times\cdots\times\Theta_k$, respectively. While

Propositions 10 and 11 enable the parameters of predicted and updated multi-object densities to be expressed analytically, the number of components grows exponentially. For the purpose of verifying the established results, we simply truncate the multi-object densities by keeping components with most significant weights.

A. Update truncation

Given a predicted density of the form (39), we approximate the updated density (49) by a truncated version:

$$\boldsymbol{\pi}(\mathbf{X}|Z) \approx \Delta(\mathbf{X}) \sum_{(I,\xi)\in\mathcal{F}(\mathbb{L})\times\Xi} \sum_{\theta\in\Theta^{(M)}} \tilde{\omega}^{(I,\xi,\theta)}(Z)\delta_I(\mathcal{L}(\mathbf{X})) \times \left[p^{(\xi,\theta)}(\cdot|Z)\right]^{\mathbf{X}},$$

where for a fixed (I,ξ), $\Theta^{(M)} = \{\theta^{(1)},...,\theta^{(M)}\}$ is the set of M elements of Θ with the highest weights $\omega^{(I,\xi,\theta^{(i)})}$, and, $\tilde{\omega}^{(I,\xi,\theta)}$ is the re-normalized weight after truncation. It is implicit that the value M can and should be chosen differently for each (I,ξ).

Enumerating the measurement set $Z = \{z_1,...,z_m\}$ and $I = \{\ell_1,...,\ell_n\}$, the M-best components $\theta^{(1)},...,\theta^{(M)}$ can be determined (without evaluating the entire set of weights) by using Murty's algorithm [32], to find the M best solution matrices $S_1,...,S_M$ to the optimal assignment problem with $(|Z|+|I|)\times(|Z|+|I|)$ cost matrix $C_Z^{(I,\xi)}$ where

$$C_Z^{(I,\xi)}(i,j) = -1(i\leq|I|)1(j\leq|Z|) \times \log\left(\frac{\left\langle p^{(\xi)}(\cdot,\ell_i), p_D(\cdot,\ell_i)g(z_j|\cdot,\ell_i)\right\rangle}{\left\langle p^{(\xi)}(\cdot,\ell_i), q_D(\cdot,\ell_i)\right\rangle\kappa(z_j)}\right).$$

Note that $C_Z^{(I,\xi)}(i,j) = 0$ if $i > |I|$ or $j > |Z|$. A solution matrix S_k consists of 0 or 1 entries with each row and column summing to 1, and a 1 in row i and column j indicates a mapping of i to j. Hence S_k represents the map $\theta^{(k)} : I \to \{0,1,...,|Z|\}$ given by

$$\theta^{(k)}(\ell_i) = \sum_{j=1}^{|Z|} j\delta(1 - S_k(i,j))$$

with the kth smallest sum $\sum_{i=1}^{|I|} C_Z^{(I,\xi)}(i,\theta^{(k)}(\ell_i))$, or equivalently kth largest product $\prod_{i=1}^{|I|}\exp(-C_Z^{(I,\xi)}(i,\theta^{(k)}(\ell_i)))$. This product is proportional to $[\eta_Z^{(\xi,\theta^{(k)})}]^I = \prod_{\ell\in I}\left\langle p^{(\xi)}(\cdot,\ell), \psi_Z(\cdot,\ell;\theta^{(k)})\right\rangle$, as can be seen from the alternative expression for $\psi_Z(\cdot,\ell;\theta^{(k)})$ in (24). Thus, $\theta^{(k)}$ is the kth best component in terms of the updated weight $\omega^{(I,\xi,\theta^{(k)})}(Z) \propto \omega^{(I,\xi)}[\eta_Z^{(\xi,\theta^{(k)})}]^I$. Since Murty's algorithm is quartic in complexity [32], the proposed approximation is quartic in the number of measurements.

B. Prediction truncation

We assume a labeled Poisson birth model (27)-(28), and use the equivalent form (58) for the prediction taken from the proof of Proposition 10:

$$\boldsymbol{\pi}_+(\mathbf{X}_+) = \Delta(\mathbf{X}_+) \sum_{(I,\xi)\in\mathcal{F}(\mathbb{L})\times\Xi} \omega^{(I,\xi)}[q_S^{(\xi)}]^I \sum_{J'\in\mathcal{F}(\mathbb{B})} w_B(J') \times \sum_{Y\in\mathcal{F}(I)} \left[\frac{\eta_S^{(\xi)}}{q_S^{(\xi)}}\right]^Y \delta_{Y\cup J'}(\mathcal{L}(\mathbf{X}_+)) \left[p_+^{(\xi)}\right]^{\mathbf{X}_+},$$

The predicted mixture $\boldsymbol{\pi}_+(\cdot)$ is approximated by a truncated version:

$$\boldsymbol{\pi}_+(\mathbf{X}_+) \approx \Delta(\mathbf{X}_+) \sum_{(I,\xi)\in\mathcal{F}(\mathbb{L})\times\Xi} \sum_{J'\in[\mathcal{F}(\mathbb{B})]^{(m')}} \sum_{Y\in[\mathcal{F}(I)]^{(m)}} \tilde{\omega}^{(I,\xi)}(J',Y) \times \delta_{Y\cup J'}(\mathcal{L}(\mathbf{X}_+)) \left[p_+^{(\xi)}\right]^{\mathbf{X}_+},$$

where for a fixed (I,ξ), $[\mathcal{F}(\mathbb{B})]^{(m')} = \{J^{(1)},...,J^{(m')}\}$ is the set of m' elements of $\mathcal{F}(\mathbb{B})$ with the highest weights $w_B(J^{(i)})$, $[\mathcal{F}(I)]^{(m)} = \{Y^{(1)},...,Y^{(m)}\}$ is the set of m elements of $\mathcal{F}(I)$ with the highest weights $\omega^{(I,\xi)}[q_S^{(\xi)}]^I[\eta_S^{(\xi)}/q_S^{(\xi)}]^{Y^{(i)}}$, and $\tilde{\omega}^{(I,\xi)}(J',Y)$ is the re-normalized weight after truncation.

For a given (I,ξ), $[\mathcal{F}(\mathbb{B})]^{(m')}$ can be trivially evaluated for a labeled Poisson RFS of births, and $[\mathcal{F}(I)]^{(m)}$ can be simply evaluated using a similar computation strategy as described in the update or via the kth shortest path algorithm. Note also that the multi-Bernoulli birth model (29)-(30) can be accommodated in an even simpler manner by treating each multi-Bernoulli component as effectively a 'survival' from 'nothing'.

C. Linear Gaussian multi-target model

For a linear Gaussian multi-target model [25], which is essentially the assumption of Gaussian mixture single target transition, likelihood and birth intensity, as well as constant survival and detection probabilities, each relevant single target density $p_{k-1}^{(\xi)}(\cdot,\ell)$ is represented as a Gaussian mixture. The corresponding Gaussian mixture predicted and updated densities $p_{k|k-1}^{(\xi)}(\cdot,\ell)$, $p_k^{(\xi)}(\cdot,\ell)$ are computed using the standard Gaussian mixture update and prediction formulas based on the Kalman filter. Associated weights, inner products, and normalizing constants can be computed from relevant Gaussian identities. This implementation is known as the para-Gaussian multi-target filter and a numerical example can be found in our preliminary work [18].

D. Non-linear multi-target model

For non-linear non-Gaussian multi-target models (with state dependent survival and detection probabilities), each single target density $p_{k-1}^{(\xi)}(\cdot,\ell)$ is represented by a set of weighted particles. The corresponding predicted and updated densities $p_{k|k-1}^{(\xi)}(\cdot,\ell)$, $p_k^{(\xi,\theta)}(\cdot,\ell)$ are computed by the standard particle (or Sequential Monte Carlo) filter. Associated weights, inner products, and normalizing constants can be computed from the particles and their weights.

E. Multi-target estimation

Given a multi-object posterior density, several solutions to multi-object state estimation such as the Joint Multi-object Estimator or Marginal Multi-object Estimator are available [1]. While these estimators are Bayes optimal, they are difficult to compute. We use a suboptimal version of the Marginal Multi-object Estimator, with the maximum *a posteriori* estimate of the cardinality and the mean estimate of the multi-object state conditioned on the estimated cardinality.

F. Numerical example

Consider a non-linear multi-target scenario with a total of 10 targets. The number of targets varies in time due to births and deaths, and the observations are subject to missed detections and clutter. Ground truths are shown in Figure 1. The target state $x_k = [\ \tilde{x}_k^T, \omega_k\ ]^T$ comprises the planar position and velocity $\tilde{x}_k = [\ p_{x,k}, \dot{p}_{x,k}, p_{y,k}, \dot{p}_{y,k}\ ]^T$ and the turn rate ω_k. Sensor returns are bearings and range vectors of the form $z_k = [\ \theta_k, r_k\ ]^T$ on the half disc of radius $2000m$.

Individual targets follow a coordinated turn model with transition density $f_{k|k-1}(x'|x_k) = \mathcal{N}(x'; m(x_k), Q)$, where $m(x_k) = [\ [F(\omega_k)\tilde{x}_k]^T, \omega_k\ ]^T$, $Q = \mathrm{diag}([\sigma_w^2 GG^T, \sigma_u^2])$,

$$F(\omega) = \begin{bmatrix} 1 & \frac{\sin\omega\Delta}{\omega} & 0 & -\frac{1-\cos\omega\Delta}{\omega} \\ 0 & \cos\omega\Delta & 0 & -\sin\omega\Delta \\ 0 & \frac{1-\cos\omega\Delta}{\omega} & 1 & \frac{\sin\omega\Delta}{\omega} \\ 0 & \sin\omega\Delta & 0 & \cos\omega\Delta \end{bmatrix}, G = \begin{bmatrix} \frac{\Delta^2}{2} & 0 \\ \Delta & 0 \\ 0 & \frac{\Delta^2}{2} \\ 0 & \Delta \end{bmatrix},$$

and $\Delta = 1s$ is the sampling time, $\sigma_w = 15m/s^2$ is the standard deviation of the process noise, $\sigma_u = \pi/180 rad/s$ is the standard deviation of the turn rate noise. The survival probability for targets is $p_{S,k}(x) = 0.99$. The birth process follows a labeled Poisson RFS with intensity $\gamma_k(x) = \sum_{i=1}^{4} w_\gamma^{(i)} \mathcal{N}(x; m_\gamma^{(i)}, P_\gamma)$ where $w_\gamma^{(1)} = w_\gamma^{(2)} = 0.02$ and $w_\gamma^{(3)} = w_\gamma^{(4)} = 0.03$, $m_\gamma^{(1)} = [\ -1500, 0, 250, 0\ \ 0\ ]^T$, $m_\gamma^{(2)} = [-250, 0, 1000, 0\ \ 0\ ]^T$, $m_\gamma^{(3)} = [\ 250, 0, 750, 0\ \ 0\ ]^T$, $m_\gamma^{(4)} = [\ 1000, 0, 1500, 0\ \ 0\ ]^T$, and $P_\gamma = \mathrm{diag}([\ 50,\ 50,\ 50,\ 50,\ 6(\pi/180)\ ]^T)^2$.

If detected, each target produces a noisy bearing and range measurement $z = [\ \theta, r\ ]^T$ with likelihood $g_k(z|x) = \mathcal{N}(z; \mu(x), R)$, where $\mu(x) = [\arctan(p_x/p_y), \sqrt{p_x^2 + p_y^2}]$ and $R = \mathrm{diag}([\ \sigma_\theta^2, \sigma_r^2\ ]^T)$ with $\sigma_\theta = (\pi/180) rad$ and $\sigma_r = 5m$. The probability of detection is state dependent and is given by $p_{D,k}(x) \propto \mathcal{N}(x; [0, 0], diag([6000, 6000])^2)$, which reaches a peak value of 0.98 at the origin and tapers to a value of 0.92 at the edge of the surveillance region. Clutter follows a Poisson RFS with intensity $\kappa_k(z) = \lambda_c \mathcal{U}(\mathbb{Z})$, where $\lambda_c = 3.2 \times 10^{-3}\ (radm)^{-1}$ and $\mathcal{U}(\mathbb{Z})$ denotes a uniform density on the observation region (giving an average of 20 clutter points per scan).

The multi-target Bayes (MTB) filter is implemented in the proposed δ-generalized labeled multi-Bernoulli form, with (continuous) kinematic component of the target state represented as weighted particle sets which are predicted and updated using standard sequential importance resampling (SIR) particle filter techniques. The labels of the target state uniquely identify target tracks. The number of new components calculated and stored in each forward propagation is set to be proportional to the weight of the original component, subject to each cardinality retaining a minimum of 33 terms, and with the entire density further truncated to a maximum of 777 terms. Components with weights below a threshold of 10^{-5} are discarded.

The filter output for a single sample run is given in Figure 2, showing the true and estimated tracks in x and y coordinates versus time. It can be seen that the filter initiates and terminates tracks with a small delay, and generally produces accurate estimates of the individual target states and the total target numbers. It is also noted that there is a small incidence of dropped or false tracks as expected, although no track switching is observed, and thus the estimated track identities are consistent throughout the entire scenario.

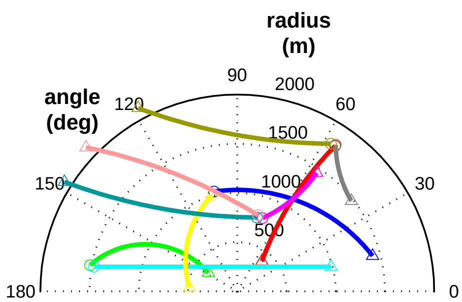

Fig. 1. Tracks in the $r\theta$ plane. Start/Stop positions for each track are shown with ○/△.

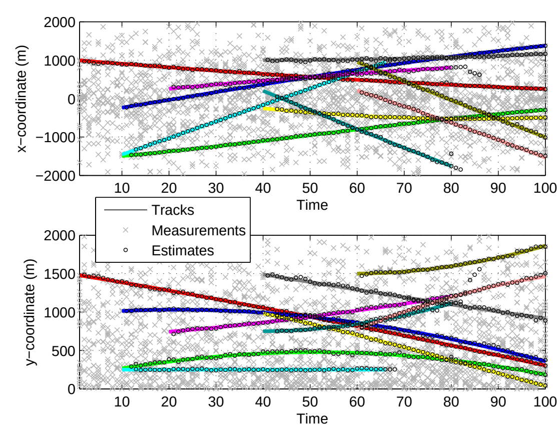

Fig. 2. Estimates and tracks for x and y coordinates versus time.

We further verify our results by comparing performance with a sequential Monte Carlo (SMC) implementation of the CPHD filter. Over 100 Monte Carlo trials, the cardinality statistics for both filters is shown in Figure 3. While both filters appear to estimate target numbers correctly, the MTB appears to outperform the CPHD on both mean and variance. Figure 4 compares the OSPA miss distance ($p = 1$, $c = 100m$) [33] for the two filters. These results confirm that, for this scenario, the MTB filter generally outperforms the CPHD filter, but incurs of an order of magnitude increase in computational complexity. The wide performance gap can be partially attributed to the SMC implementation of the CPHD filter which is error prone due its reliance on clustering to extract state estimates (whereas the MTB filter does not require clustering). We stress that the aim of our comparison is to verify the established theoretical results and is not intended to be a comprehensive study. A rigorous performance analysis of the proposed algorithm is beyond the scope of the paper and is the subject of further work.

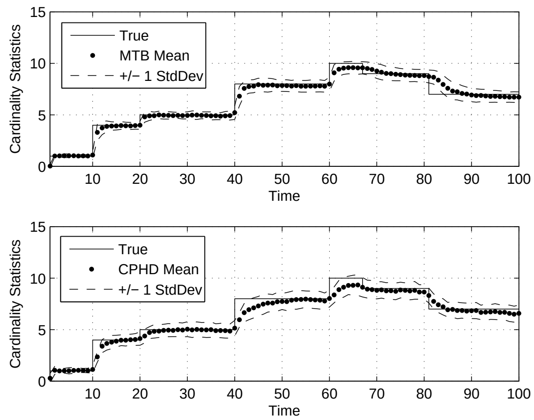

Fig. 3. Cardinality statistics for the MTB and CPHD filters versus time.

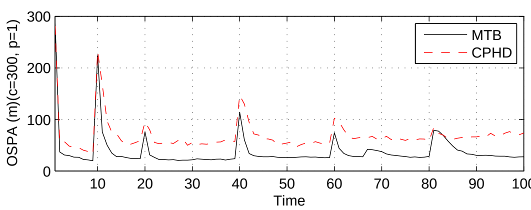

Fig. 4. OSPA distance ($p = 1$, $c = 300m$) for MTB and CPHD filters.

VII. Conclusion

By introducing the notion of labeled RFSs, we were able to construct two families of multi-object conjugate priors that are also closed under the multi-object Chapman-Kolmogorov equation. A multi-target tracking filter, completely described by the multi-object prediction and update equations, is proposed to demonstrate the application of these results. To the best of our knowledge, this is the first RFS-based multi-target filter that produces track-valued estimate in a principled manner, thereby refuting the opinion that the RFS approach alone cannot produce track estimates. Our results can potentially be extended to multi-object smoothing. Implementation of the algorithm employs Murty's algorithm to reduce the number of components in the multi-object densities. This complexity can be further reduced by optimizing Murty's algorithm [34] or by using Lagrangian relaxation [35].

Appendix

The following Lemma is used in several instances throughout the Appendix.

Lemma 12 *If $f : \mathbb{L}^n \to \mathbb{R}$ is symmetric, i.e. its value at any n-tuple of arguments is the same as its value at any permutation of that n-tuple, then*

$$\sum_{(\ell_1,\dots,\ell_n)\in\mathbb{L}^n} \delta_n(|\{\ell_1,\dots,\ell_n\}|) f(\ell_1,\dots,\ell_n) = n! \sum_{\{\ell_1,\dots,\ell_n\}\in\mathcal{F}_n(\mathbb{L})} f(\ell_1,\dots,\ell_n)$$

Proof: Due to the term $\delta_n(|\{\ell_1,\dots,\ell_n\}|)$, the sum becomes a sum over indices in $\mathbb{L}^n$ with distinct components. The symmetry of f means that all $n!$ permutations of $(\ell_1,\dots,\ell_n)$ shares the same functional value $f(\ell_1,\dots,\ell_n)$. Moreover, all $n!$ permutations of $(\ell_1,\dots,\ell_n)$ with distinct components define an equivalence class $\{\ell_1,\dots,\ell_n\} \in \mathcal{F}_n(\mathbb{L})$. □

Proof of Lemma 3:

$$\begin{aligned}
&\int \delta_{|\mathbf{X}|}(|\mathcal{L}(\mathbf{X})|) h(\mathcal{L}(\mathbf{X})) g^{\mathbf{X}} \delta\mathbf{X} \\
&= \sum_{n=0}^{\infty} \frac{1}{n!} \sum_{(\ell_1,\dots\ell_n)\in\mathbb{L}^n} \delta_n(|\{\ell_1,\dots,\ell_{n'}\}|) h(\{\ell_1,\dots,\ell_n\}) \\
&\qquad \times \int \left(\prod_{i=1}^{n} g(x_i,\ell_i) \right) dx_1 \dots dx_n \\
&= \sum_{n=0}^{\infty} \frac{1}{n!} \sum_{(\ell_1,\dots\ell_n)\in\mathbb{L}^n} \delta_n(|\{\ell_1,\dots,\ell_n\}|) h(\{\ell_1,\dots,\ell_n\}) \\
&\qquad \times \prod_{i=1}^{n} \left(\int g(x_i,\ell_i) dx_i \right) \\
&= \sum_{n=0}^{\infty} \sum_{\{\ell_1,\dots\ell_n\}\in\mathcal{F}_n(\mathbb{L})} h(\{\ell_1,\dots\ell_n\}) \prod_{i=1}^{n} \left(\int g(x_i,\ell_i) dx_i \right),
\end{aligned}$$

where the last line follows from the symmetry of $h(\{\ell_1,\dots\ell_n\}) \prod_{i=1}^{n} \left(\int g(x_i,\ell_i) dx_i \right)$ in $(\ell_1,\dots\ell_n)$, and Lemma 12. The double sum can be combined as a sum over the subsets of $\mathbb{L}$ and the result follows. □

Proof of Proposition 5: From [1], for any RFS $\mathbf{X}$ on $\mathbb{X}\times\mathbb{L}$ the cardinality distribution is given by

$$\boldsymbol{\rho}(n) = \frac{1}{n!} \int_{(\mathbb{X}\times\mathbb{L})^n} \boldsymbol{\pi}(\{\mathbf{x}_1,\dots,\mathbf{x}_n\}) d(\mathbf{x}_1\dots,\mathbf{x}_n)$$

Hence, substituting for $\boldsymbol{\pi}$ using the defining equation for generalized labeled multi-Bernoulli (13) and then bringing the integral inside the sum over $\mathbb{C}$ give

$$\begin{aligned}
&\boldsymbol{\rho}(n) \\
&= \frac{1}{n!} \sum_{(\ell_1,\dots,\ell_n)\in\mathbb{L}^n} \delta_n(|\{\ell_1,\dots,\ell_n\}|) \sum_{c\in\mathbb{C}} w^{(c)}(\{\ell_1,\dots,\ell_n\}) \\
&\qquad \times \left[\int_{\mathbb{X}^n} \left(\prod_{i=1}^{n} p^{(c)}(x_i,\ell_i) \right) d(x_1,\dots,x_n) \right] \\
&= \sum_{c\in\mathbb{C}} \frac{1}{n!} \sum_{(\ell_1,\dots,\ell_n)\in\mathbb{L}^n} \delta_n(|\{\ell_1,\dots,\ell_n\}|) w^{(c)}(\{\ell_1,\dots,\ell_n\}) \\
&\qquad \times \left(\prod_{i=1}^{n} \int_{\mathbb{X}} p^{(c)}(x_i,\ell_i) dx_i \right)
\end{aligned}$$

$$\begin{aligned} &= \sum_{c\in\mathbb{C}} \frac{1}{n!} \sum_{(\ell_1,...,\ell_n)\in\mathbb{L}^n} \delta_n(|\{\ell_1,...,\ell_n\}|) w^{(c)}(\{\ell_1,...,\ell_n\}) \\ &= \sum_{c\in\mathbb{C}} \frac{1}{n!} n! \sum_{\{\ell_1,...,\ell_n\}\in\mathcal{F}_n(\mathbb{L})} w^{(c)}(\{\ell_1,...,\ell_n\}) \quad \text{(using Lemma 12)} \\ &= \sum_{L\in\mathcal{F}_n(\mathbb{L})} \sum_{c\in\mathbb{C}} w^{(c)}(L). \ \square \end{aligned}$$

Proof of Proposition 6: Following [1], the PHD of the unlabeled RFS is given by

$$v(x) = \sum_{n=0}^{\infty} \frac{1}{n!} \int \pi(\{x\}\cup\{x_1,...,x_n\})\, d(x_1...,x_n)$$

where

$$\pi(\{x\}\cup\{x_1,...,x_n\}) = \sum_{(\ell,\ell_1,...,\ell_n)\in\mathbb{L}^{n+1}} \boldsymbol{\pi}(\{(x,\ell),(x_1,\ell_1),...,(x_n,\ell_n)\})$$

by virtue of (9). Similar to the proof of the cardinality distribution, substituting for $\boldsymbol{\pi}$ using the defining equation for generalized labeled multi-Bernoulli (13), then bringing the integral inside the sum over $\mathbb{C}$ and using the property that each $p^{(c)}(\cdot,\ell_i)$ integrates to 1 gives

$$\begin{aligned} v(x) &= \sum_{n=0}^{\infty} \frac{1}{n!} \sum_{c\in\mathbb{C}} \sum_{(\ell,\ell_1,...,\ell_n)\in\mathbb{L}^{n+1}} \delta_{n+1}(|\{\ell,\ell_1,...,\ell_n\}|) \\ &\qquad \times w^{(c)}(\{\ell,\ell_1,...,\ell_n\}) p^{(c)}(x,\ell) \\ &= \sum_{n=0}^{\infty} \frac{1}{n!} \sum_{c\in\mathbb{C}} \sum_{\ell\in\mathbb{L}^n} \sum_{(\ell_1,...,\ell_n)\in\mathbb{L}^n} \delta_n(|\{\ell_1,...,\ell_n\}|) \\ &\qquad \times \left(1 - 1_{\{\ell_1,...,\ell_n\}}(\ell)\right) w^{(c)}(\{\ell,\ell_1,...,\ell_n\}) p^{(c)}(x,\ell) \end{aligned}$$

Since $\left(1 - 1_{\{\ell_1,...,\ell_n\}}(\ell)\right) w^{(c)}(\{\ell,\ell_1,...,\ell_n\})$ is permutation invariant in $(\ell_1,...,\ell_n)$, applying Lemma 12 gives

$$\begin{aligned} v(x) &= \sum_{n=0}^{\infty} \sum_{c\in\mathbb{C}} \sum_{\ell\in\mathbb{L}} p^{(c)}(x,\ell) \sum_{L\in\mathcal{F}_n(\mathbb{L})} (1 - 1_L(\ell))\, w^{(c)}(\{\ell\}\cup L) \\ &= \sum_{c\in\mathbb{C}} \sum_{\ell\in\mathbb{L}} p^{(c)}(x,\ell) \sum_{L\subseteq\mathbb{L}} (1 - 1_L(\ell))\, w^{(c)}(\{\ell\}\cup L) \\ &= \sum_{c\in\mathbb{C}} \sum_{\ell\in\mathbb{L}} p^{(c)}(x,\ell) \sum_{I\subseteq\mathbb{L}} 1_I(\ell) w^{(c)}(I). \square \end{aligned}$$

Proof of Proposition 7: Following [1] pp. 410, the multi-object likelihood (18) can be equivalently rewritten in the following form

$$g(Z|\mathbf{X}) = e^{-\langle\kappa,1\rangle}\kappa^Z \sum_{\theta\in\Theta} \delta_{\theta^{-1}(\{0:|Z|\})}(\mathcal{L}(\mathbf{X}))[\psi_Z(\cdot;\theta)]^{\mathbf{X}} \quad (54)$$

where Θ is the space of mappings $\theta:\mathbb{L}\to\{0:|Z|\}\triangleq\{0,1,...,|Z|\}$, such that $\theta(i)=\theta(i')>0$ implies $i=i'$, i.e when the range is restricted to the positive integers, θ is one-to-one. Note that the multi-object likelihood in [1] involves summing $[\psi_Z(\cdot;\theta)]^{\mathbf{X}}$ over the θ's that are defined on $\{1:|\mathbf{X}|\}$ (instead of $\mathbb{L}$ as per our definition). Nonetheless, this is equivalent to (54) since the term $\delta_{\theta^{-1}(\{0:|Z|\})}(\mathcal{L}(\mathbf{X}))$ restricts the sum to the θ's with domain $\mathcal{L}(\mathbf{X})$. For convenience abbreviate $\lambda = e^{-\langle\kappa,1\rangle}\kappa^Z$, then

$$\begin{aligned} g(Z|\mathbf{X})\boldsymbol{\pi}(\mathbf{X}) &= \Delta(\mathbf{X})\lambda \sum_{c\in\mathbb{C}} \sum_{\theta\in\Theta} \delta_{\theta^{-1}(\{0:|Z|\})}(\mathcal{L}(\mathbf{X})) \\ &\qquad \times w^{(c)}(\mathcal{L}(\mathbf{X}))[p^{(c)}\psi_Z(\cdot;\theta)]^{\mathbf{X}} \\ &= \Delta(\mathbf{X})\lambda \sum_{c\in\mathbb{C}} \sum_{\theta\in\Theta} \delta_{\theta^{-1}(\{0:|Z|\})}(\mathcal{L}(\mathbf{X})) \\ &\qquad \times w^{(c)}(\mathcal{L}(\mathbf{X}))[\eta_Z^{(c,\theta)}]^{\mathcal{L}(\mathbf{X})} \left[p^{(c,\theta)}(\cdot|Z)\right]^{\mathbf{X}} \end{aligned}$$

since $p^{(c)}(x,\ell)\psi_Z(x,\ell;\theta) = \eta_Z^{(c,\theta)}(\ell)p^{(c,\theta)}(x,\ell|Z)$ by virtue of (21). Now

$$\begin{aligned} &\int g(Z|\mathbf{X})\boldsymbol{\pi}(\mathbf{X})\delta\mathbf{X} \\ &= \lambda \int \Delta(\mathbf{X}) \sum_{c\in\mathbb{C}} \sum_{\theta\in\Theta} \delta_{\theta^{-1}(\{0:|Z|\})}(\mathcal{L}(\mathbf{X})) \\ &\qquad \times w^{(c)}(\mathcal{L}(\mathbf{X}))[\eta_Z^{(c,\theta)}]^{\mathcal{L}(\mathbf{X})} \left[p^{(c,\theta)}(\cdot|Z)\right]^{\mathbf{X}} \delta\mathbf{X} \\ &= \lambda \sum_{c\in\mathbb{C}} \sum_{\theta\in\Theta} \int \Delta(\mathbf{X}) \delta_{\theta^{-1}(\{0:|Z|\})}(\mathcal{L}(\mathbf{X})) \\ &\qquad \times w^{(c)}(\mathcal{L}(\mathbf{X}))[\eta_Z^{(c,\theta)}]^{\mathcal{L}(\mathbf{X})} \left[p^{(c,\theta)}(\cdot|Z)\right]^{\mathbf{X}} \delta\mathbf{X} \\ &= \lambda \sum_{c\in\mathbb{C}} \sum_{\theta\in\Theta} \sum_{J\subseteq\mathbb{L}} \delta_{\theta^{-1}(\{0:|Z|\})}(J) w^{(c)}(J) [\eta_Z^{(c,\theta)}]^J \end{aligned}$$

where the last line follows from Lemma 3. Hence,

$$\begin{aligned} &\boldsymbol{\pi}(\mathbf{X}|Z) \\ &= \frac{g(Z|\mathbf{X})\boldsymbol{\pi}(\mathbf{X})\delta\mathbf{X}}{\int g(Z|\mathbf{X})\boldsymbol{\pi}(\mathbf{X})\delta\mathbf{X}} \\ &= \frac{\Delta(\mathbf{X}) \sum_{c\in\mathbb{C}} \sum_{\theta} \left[\delta_{\theta^{-1}(\{0:|Z|\})}(\mathcal{L}(\mathbf{X})) w^{(c)}(\mathcal{L}(\mathbf{X}))[\eta_Z^{(c,\theta)}]^{\mathcal{L}(\mathbf{X})}\right]}{\sum_{c\in\mathbb{C}} \sum_{\theta} \sum_{J\subseteq\mathbb{L}} \delta_{\theta^{-1}(\{0:|Z|\})}(J) w^{(c)}(I) [\eta_Z^{(c,\theta)}]^J} \\ &\quad \times \left[p^{(c,\theta)}(\cdot|Z)\right]^{\mathbf{X}} \\ &= \Delta(\mathbf{X}) \sum_{c\in\mathbb{C}} \sum_{\theta} w_Z^{(c,\theta)}(\mathcal{L}(\mathbf{X})) \left[p^{(c,\theta)}(\cdot|Z)\right]^{\mathbf{X}} \quad \text{(from (22))}.\square \end{aligned}$$

Proof of Proposition 8: The density of the surviving multi-object state at the next time is given by the Chapman-Kolmogorov equation

$$\begin{aligned} &\boldsymbol{\pi}_S(\mathbf{W}) \\ &= \int \mathbf{f}_S(\mathbf{W}|\mathbf{X})\boldsymbol{\pi}(\mathbf{X})\delta\mathbf{X}. \\ &= \Delta(\mathbf{W}) \int 1_{\mathcal{L}(\mathbf{X})}(\mathcal{L}(\mathbf{W}))[\Phi(\mathbf{W};\cdot)]^{\mathbf{X}} \\ &\qquad \times \Delta(\mathbf{X}) \sum_{c\in\mathbb{C}} w^{(c)}(\mathcal{L}(\mathbf{X})) \left[p^{(c)}\right]^{\mathbf{X}} \delta\mathbf{X} \\ &= \Delta(\mathbf{W}) \sum_{c\in\mathbb{C}} \int \Delta(\mathbf{X}) 1_{\mathcal{L}(\mathbf{X})}(\mathcal{L}(\mathbf{W})) w^{(c)}(\mathcal{L}(\mathbf{X})) \\ &\qquad \times \left[\Phi(\mathbf{W};\cdot)p^{(c)}\right]^{\mathbf{X}} \delta\mathbf{X} \end{aligned}$$

$$= \Delta(\mathbf{W})\sum_{c\in\mathbb{C}}\sum_{I\subseteq\mathbb{L}} 1_I(\mathcal{L}(\mathbf{W}))w^{(c)}(L)\prod_{\ell\in I}\left\langle\Phi(\mathbf{W};\cdot,\ell),p^{(c)}(\cdot,\ell)\right\rangle \tag{55}$$

where (55) follows from the previous line via Lemma 3. Due to the term $1_I(\mathcal{L}(\mathbf{W}))$, we only need to consider $I\supseteq\mathcal{L}(\mathbf{W})$, in which case

$$\begin{aligned}
&\prod_{\ell\in I}\left\langle\Phi(\mathbf{W};\cdot,\ell),p^{(c)}(\cdot,\ell)\right\rangle\\
&= \prod_{\ell\in\mathcal{L}(\mathbf{W})}\left\langle\Phi(\mathbf{W};\cdot,\ell),p^{(c)}(\cdot,\ell)\right\rangle\prod_{\ell\in I-\mathcal{L}(\mathbf{W})}\left\langle\Phi(\mathbf{W};\cdot,\ell),p^{(c)}(\cdot,\ell)\right\rangle\\
&= \prod_{\ell\in\mathcal{L}(\mathbf{W})}\sum_{(x_+,\ell_+)\in\mathbf{W}}\delta_\ell(\ell_+)\left\langle p_S(\cdot,\ell)f(x_+|\cdot,\ell),p^{(c)}(\cdot,\ell)\right\rangle\\
&\qquad\times\prod_{\ell\in I-\mathcal{L}(\mathbf{W})}\left\langle q_S(\cdot,\ell),p^{(c)}(\cdot,\ell)\right\rangle\\
&= \prod_{\ell\in\mathcal{L}(\mathbf{W})}\sum_{(x_+,\ell_+)\in\mathbf{W}}\delta_\ell(\ell_+)p_S^{(c)}(x_+,\ell)\eta_S^{(c)}(\ell)\prod_{\ell\in I-\mathcal{L}(\mathbf{W})}q_S^{(c)}(\ell)\\
&= \prod_{(x_+,\ell)\in\mathbf{W}}p_S^{(c)}(x_+,\ell)\eta_S^{(c)}(\ell)\prod_{\ell\in I-\mathcal{L}(\mathbf{W})}q_S^{(c)}(\ell)\\
&= \left[p_S^{(c)}\right]^{\mathbf{W}}\left[\eta_S^{(c)}\right]^{\mathcal{L}(\mathbf{W})}\left[q_S^{(c)}\right]^{I-\mathcal{L}(\mathbf{W})}
\end{aligned}$$

Hence, substituting this into (55) and using (37) for $w_S^{(c)}(L)$ gives

$$\boldsymbol{\pi}_S(\mathbf{W}) = \Delta(\mathbf{W})\sum_{c\in\mathbb{C}}w_S^{(c)}(\mathcal{L}(\mathbf{W}))[p_S^{(c)}]^{\mathbf{W}}$$

For the predicted multi-object density, recall the birth density (26), let $\mathbf{X}_B=\mathbf{X}_+-\mathbb{X}\times\mathbb{L}$ and $\mathbf{X}_S=\mathbf{X}_+\cap\mathbb{X}\times\mathbb{L}$, then

$$\begin{aligned}
&\boldsymbol{\pi}_+(\mathbf{X}_+)\\
&= \int\mathbf{f}(\mathbf{X}_+|\mathbf{X})\boldsymbol{\pi}(\mathbf{X})\delta\mathbf{X}\\
&= \mathbf{f}_B(\mathbf{X}_B)\int\mathbf{f}_S(\mathbf{X}_S|\mathbf{X})\boldsymbol{\pi}(\mathbf{X})\delta\mathbf{X}\\
&= \mathbf{f}_B(\mathbf{X}_B)\boldsymbol{\pi}_S(\mathbf{X}_S)\\
&= \Delta(\mathbf{X}_B)\Delta(\mathbf{X}_S)\sum_{c\in\mathbb{C}}w_B(\mathcal{L}(\mathbf{X}_B))w_S^{(c)}(\mathcal{L}(\mathbf{X}_S))[p_B]^{\mathbf{X}_B}[p_S^{(c)}]^{\mathbf{X}_S}\\
&= \Delta(\mathbf{X}_+)\sum_{c\in\mathbb{C}}w_B(\mathcal{L}(\mathbf{X}_+)-\mathbb{L}))w_S^{(c)}(\mathcal{L}(\mathbf{X}_+)\cap\mathbb{L}))[p_+^{(c)}]^{\mathbf{X}_+}\\
&\qquad\qquad\text{(since }\mathcal{L}\text{ is a projection)}\\
&= \Delta(\mathbf{X}_+)\sum_{c\in\mathbb{C}}w_+^{(c)}(\mathcal{L}(\mathbf{X}_+))[p_+^{(c)}]^{\mathbf{X}_+}.\ \square
\end{aligned}$$

Proof of Proposition 10: From Proposition 8,

$$\boldsymbol{\pi}_+(\mathbf{X}_+) = \Delta(\mathbf{X}_+)\sum_{(I,\xi)\in\mathcal{F}(\mathbb{L})\times\Xi}w_+^{(I,\xi)}(\mathcal{L}(\mathbf{X}_+))\left[p_+^{(\xi)}\right]^{\mathbf{X}_+}$$

where

$$w_+^{(I,\xi)}(L) = w_B(L-\mathbb{L})w_S^{(I,\xi)}(L\cap\mathbb{L}) \tag{56}$$

$$w_S^{(I,\xi)}(J) = [\eta_S^{(\xi)}]^J\sum_{L\subseteq\mathbb{L}}1_L(J)\delta_I(L)\omega^{(I,\xi)}[q_S^{(\xi)}]^{L-J} \tag{57}$$

Note that $1_L(J)\delta_I(L)\omega^{(I,\xi)}[q_S^{(\xi)}]^L = 1_I(J)\omega^{(I,\xi)}[q_S^{(\xi)}]^I$ if $L=I$ and zero otherwise, hence

$$\begin{aligned}
w_S^{(I,\xi)}(J) &= [\eta_S^{(\xi)}]^J 1_I(J)\omega^{(I,\xi)}[q_S^{(\xi)}]^{I-J}\\
&= [\eta_S^{(\xi)}]^J\sum_{Y\subseteq I}\delta_Y(J)\omega^{(I,\xi)}[q_S^{(\xi)}]^{I-J}\\
&= \sum_{Y\subseteq I}\delta_Y(J)[\eta_S^{(\xi)}]^J\omega^{(I,\xi)}[q_S^{(\xi)}]^{I-J}\\
&= \sum_{Y\subseteq I}\delta_Y(J)[\eta_S^{(\xi)}]^Y\omega^{(I,\xi)}[q_S^{(\xi)}]^{I-Y}
\end{aligned}$$

Rewriting $w_B(L)=\sum_{J'\in\mathcal{F}(\mathbb{B})}w_B(J')\delta_{J'}(L)$ and substitute this into (56), we have

$$\begin{aligned}
&w_+^{(I,\xi)}(L)\\
&= \sum_{Y\subseteq I}\delta_Y(L\cap\mathbb{L})[\eta_S^{(\xi)}]^Y\omega^{(I,\xi)}[q_S^{(\xi)}]^{I-Y}\sum_{J'\in\mathcal{F}(\mathbb{B})}w_B(J')\delta_{J'}(L-\mathbb{L})\\
&= \omega^{(I,\xi)}\sum_{J'\in\mathcal{F}(\mathbb{B})}w_B(J')\sum_{Y\subseteq I}[\eta_S^{(\xi)}]^Y[q_S^{(\xi)}]^{I-Y}\delta_Y(L\cap\mathbb{L})\delta_{J'}(L-\mathbb{L})\\
&= \omega^{(I,\xi)}\sum_{J'\in\mathcal{F}(\mathbb{B})}w_B(J')\sum_{Y\subseteq I}[\eta_S^{(\xi)}]^Y[q_S^{(\xi)}]^{I-Y}\delta_{Y\cup J'}(L)
\end{aligned}$$

where the last line follows from $L\cap\mathbb{L}=Y$ and $L-\mathbb{L}=J'$ $\Rightarrow L=Y\cup J'$. Now

$$\begin{aligned}
&\boldsymbol{\pi}_+(\mathbf{X}_+)\\
&= \Delta(\mathbf{X}_+)\sum_{I\in\mathcal{F}(\mathbb{L})}\sum_{\xi\in\Xi}\omega^{(I,\xi)}\sum_{J'\in\mathcal{F}(\mathbb{B})}\sum_{Y\subseteq I}\delta_{Y\cup J'}(\mathcal{L}(\mathbf{X}_+))\\
&\qquad\times w_B(J')[\eta_S^{(\xi)}]^Y[q_S^{(\xi)}]^{I-Y}\left[p_+^{(\xi)}\right]^{\mathbf{X}_+} \qquad (58)\\
&= \Delta(\mathbf{X}_+)\sum_{I\in\mathcal{F}(\mathbb{L})}\sum_{\xi\in\Xi}\omega^{(I,\xi)}\sum_{J'\in\mathcal{F}(\mathbb{B})}\sum_{Y\in\mathcal{F}(\mathbb{L})}1_I(Y)\delta_{Y\cup J'}(\mathcal{L}(\mathbf{X}_+))\\
&\qquad\times w_B(J')[\eta_S^{(\xi)}]^Y[q_S^{(\xi)}]^{I-Y}\left[p_+^{(\xi)}\right]^{\mathbf{X}_+}\\
&= \Delta(\mathbf{X}_+)\sum_{Y\in\mathcal{F}(\mathbb{L})}\sum_{J'\in\mathcal{F}(\mathbb{B})}\sum_{\xi\in\Xi}\delta_{Y\cup J'}(\mathcal{L}(\mathbf{X}_+))\left[p_+^{(\xi)}\right]^{\mathbf{X}_+}\\
&\qquad\times\left(w_B(J')[\eta_S^{(\xi)}]^Y\sum_{I\in\mathcal{F}(\mathbb{L})}1_I(Y)[q_S^{(\xi)}]^{I-Y}\omega^{(I,\xi)}\right)\\
&= \Delta(\mathbf{X}_+)\sum_{Y\in\mathcal{F}(\mathbb{L})}\sum_{J'\in\mathcal{F}(\mathbb{B})}\sum_{\xi\in\Xi}\delta_{Y\cup J'}(\mathcal{L}(\mathbf{X}_+))\left[p_+^{(\xi)}\right]^{\mathbf{X}_+}\\
&\qquad\times\left(w_B(J')\omega_S^{(\xi)}(Y)\right)\\
&= \Delta(\mathbf{X}_+)\sum_{(I_+,\xi)\in\mathcal{F}(\mathbb{L}_+)\times\Xi}\delta_{I_+}(\mathcal{L}(\mathbf{X}_+))\left[p_+^{(\xi)}\right]^{\mathbf{X}_+}\\
&\qquad\times\left(w_B(I_+\cap\mathbb{B})\omega_S^{(\xi)}(I_+\cap\mathbb{L})\right)
\end{aligned}$$

where the last line follows by setting $I_+=Y\cup J'$, and since $Y\subseteq\mathbb{L}$, $J'\subset\mathbb{B}$ and $\mathbb{L}$, $\mathbb{B}$ are disjoint, $Y=I_+\cap\mathbb{L}$ and $J'=I_+\cap\mathbb{B}$. $\square$

Proof of Proposition 11: From proposition 7

$$\boldsymbol{\pi}(\mathbf{X}|Z) = \Delta(\mathbf{X})\sum_{(I,\xi)\in\mathcal{F}(\mathbb{L})\times\Xi}\sum_{\theta\in\Theta}w^{(I,\xi,\theta)}(\mathcal{L}(\mathbf{X})|Z)\left[p^{(\xi,\theta)}(\cdot|Z)\right]^{\mathbf{X}}$$

where

$$w^{(I,\xi,\theta)}(L|Z)$$
$$= \frac{\delta_{\theta^{-1}(\{0:|Z|\})}(L)\omega^{(I,\xi)}\delta_I(L)[\eta_Z^{(\xi,\theta)}]^L}{\sum_{(I,\xi)\in\mathcal{F}(\mathbb{L})\times\Xi}\sum_{\theta\in\Theta}\sum_{J\subseteq\mathbb{L}}\delta_{\theta^{-1}(\{0:|Z|\})}(J)\omega^{(I,\xi)}\delta_I(J)[\eta_Z^{(\xi,\theta)}]^J}$$

with θ, $p^{(\xi,\theta)}(x,\ell|Z)$, $\eta_Z^{(\xi,\theta)}(\ell)$, $\psi_Z(\cdot,\ell;\theta)$ given as in Proposition 11. For $w^{(I,\xi,\theta)}(L|Z)$, note that $\delta_{\theta^{-1}(\{0:|Z|\})}(L)\delta_I(L) = \delta_{\theta^{-1}(\{0:|Z|\})}(I)\delta_I(L)$ and consider the sum over $I\subseteq\mathbb{L}$ in the denominator

$$\sum_{J\subseteq\mathbb{L}}\delta_{\theta^{-1}(\{0:|Z|\})}(J)\omega^{(I,\xi)}\delta_I(J)[\eta_Z^{(\xi,\theta)}]^J$$
$$= \delta_{\theta^{-1}(\{0:|Z|\})}(I)\omega^{(I,\xi)}[\eta_Z^{(\xi,\theta)}]^I$$

we can rewrite

$$\begin{aligned} & w^{(I,\xi,\theta)}(L|Z) \\ = \; & \frac{\delta_{\theta^{-1}(\{0:|Z|\})}(I)\omega^{(I,\xi)}[\eta_Z^{(\xi,\theta)}]^I}{\sum_{(I,\xi)\in\mathcal{F}(\mathbb{L})\times\Xi}\sum_{\theta\in\Theta}\delta_{\theta^{-1}(\{0:|Z|\})}(I)\omega^{(I,\xi)}[\eta_Z^{(\xi,\theta)}]^I}\delta_I(L) \\ = \; & \omega^{(I,\xi,\theta)}(Z)\delta_I(L).\square \end{aligned}$$

REFERENCES

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

[2] A.J. Baddeley, and M.N. van Lieshout, "ICM for object recognition". In Y. Dodge and J. Whittaker (eds.), *Computational Statistics*, Vol. 2. Heidelberg: Physica/Springer, pp. 271-286, 1992.

[3] J. Mullane, B.-N. Vo, M. Adams, and B.-T. Vo, "A Random Finite Set Approach to Bayesian SLAM," *IEEE Trans. Robotics*, Vol. 27, No. 2, pp. 268-282, 2011.

[4] J. Mullane, B.-N. Vo, M. Adams and B.-T. Vo, *Random Finite Set in Robotic Map Building and SLAM*, Springer, 2011.

[5] K. Dralle and M. Rudemo, "Automatic estimation of individual tree positions from aerial photos," *Can. J. Forest Res.*, vol. 27, pp. 1728–1736, 1997.

[6] J. Lund and M. Rudemo, "Models for point processes observed with noise,"*Biometrika*, vol. 87, no. 2, pp. 235–249, 2000.

[7] J. Moller and R. Waagepeterson, "Modern Statistics for Spatial Point Processes", *Scandinavian Journal of Statistics*, Vol. 34, pp. 643–684, 2006.

[8] L.A. Waller and C.A. Gotway, *Applied Spatial Statistics for Public Health Data*, John Wiley & Sons, 2004.

[9] N.A.C. Cressie, *Statistics for Spatial Data*, John Wiley & Sons, 1993.

[10] P.J. Diggle, *Statistical Analysis of Spatial Point Patterns*, 2nd edition, Arnold Publishers, 2003.

[11] Y. Bar-Shalom and T. Fortmann, *Tracking and Data Association*, Academic Press, San Diego, 1988.

[12] C.P. Robert, *The Bayesian Choice*, Springer-Verlag, New York, 1994.

[13] C.P. Robert, and G. Casella, *Monte Carlo Statistical Methods*, Springer-Verlag, New York, 1999.

[14] H. Raiffa and R. Schlaifer, "Applied Statistical Decision Theory," Division of Research, Graduate School of Business Administration, Harvard University, 1961.

[15] A. Gelman, J.B. Carlin, H.S. Stern, and D.B. Rubin, *Bayesian Data Analysis*, 2nd edition. CRC Press, 2003.

[16] Y.W. Teh, M.I. Jordan, M.J. Beal, and D.M. Blei, "Hierarchical Dirichlet Processes," *Journal of the American Statistical Association*, Vol. 101, No. 476, pp. 1566-1581, 2006.

[17] B.-N. Vo, B.-T. Vo, N.-T. Pham and D. Suter, "Joint Detection and Estimation of Multiple Objects from Image Observations," *IEEE Trans. Signal Procesing*, Vol. 58, No. 10, pp. 5129–5241, 2010.

[18] B.-T. Vo, and B.-N. Vo, "A Random Finite Set Conjugate Prior and Application to Multi-Target Tracking," Proc. 7th Int. *Conf. Intelligent Sensors, Sensor Networks & Information Processing* (ISSNIP'2011), Adelaide, Australia, December 2011.

[19] J. Williams, "Experiments with graphical model implementations of multiple target multiple Bernoulli filters," Proc. 7th Int. *Conf. Int. Sensors, Sensor Networks & Inf. Proc.* (ISSNIP'2011), Adelaide, Australia, December 2011.

[20] R. Mahler, "Multi-target Bayes filtering via first-order multi-target moments,"*IEEE Transactions of Aerospace and Electronic Systems*, Vol. 39, No. 4, pp. 1152-1178, 2003.

[21] R. Mahler, "PHD filters of higher order in target number," *IEEE Trans. Aerospace & Electronic Systems*, Vol. 43, No. 3, pp. 1523–1543 2007.

[22] B.-T. Vo, B.-N. Vo, and A. Cantoni, "The cardinality balanced multi-target multi-Bernoulli filter and its implementations," *IEEE Trans. Signal Processing*, Vol. 57, No. 2, pp. 409–423, 2009.

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

[24] N. Whiteley, S. Singh; S. Godsill, "Auxiliary Particle Implementation of Probability Hypothesis Density Filter,", *IEEE Trans. Aerospace and Electronic Systems*, vol.46, no.3, pp.1437-1454, July 2010

[25] B.-N Vo and W.-K. Ma, "The Gaussian mixture probability hypothesis density filter", *IEEE Trans. Signal Processing*, vol. 54, no. 11, pp. 4091–4104, 2006.

[26] B.-T. Vo, B.-N. Vo, and A. Cantoni, "Analytic implementations of the cardinalized probability hypothesis density filter," *IEEE Trans. Signal Processing*, Vol. 55, No. 7, pp. 3553–3567, 2007.

[27] D. Clark and J. Bell, "Convergence results for the Particle-PHD filter," *IEEE Trans. Signal Processing*,vol. 54, no. 7, pp. 2652—2661, 2006.

[28] A. Johansen, S. Singh, A. Doucet, and B. Vo, "Convergence of the sequential Monte Carlo implementation of the PHD filter," *Methodology and Computing in Applied Probability*, vol. 8, no. 2, pp. 265-291, 2006.

[29] D. Clark, and B.-N. Vo, "Convergence analysis of the Gaussian mixture Probability Hypothesis Density filter," *IEEE Trans. Signal Processing*, Vol. 55, No. 4, pp. 1204-1212, 2007.

[30] F. Caron, P. Del Moral, M. Pace and B.-N. Vo, "On the stability and the approximation of branching distribution flows, with applications to nonlinear multiple target filtering," *Jnl. Stoc. Anal. & Appl.*, Vol. 29, No. 6, pp. 951-997, 2011.

[31] D.J. Daley and D. Vere-Jones, *An Introduction to the Theory of Point Processes*, Springer, New York, 1988.

[32] K. G. Murty, "An Algorithm for Ranking all the Assignments in Order of Increasing Cost," *Operations Research*, Vol. 16, No. 3. (1968), pp.682-687.

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

[34] M. Miller, H. Stone, and I. Cox, "Optimizing Murty's ranked assignment method," *IEEE Trans. Aerospace & Electronic Systems*, Vol. 33, No. 3, pp. 851–862, 1997.

[35] A. B. Poore and N. Rijavec, "A Lagrangian Relaxation Algorithm for Multidimensional Assignment Problems Arising from Multitarget Tracking," SIAM Journal of Optimization, Vol. 3, No. 3, pp. 544–563, 1993.