A Multi-Scan Labeled Random Finite Set Model for Multi-Object State Estimation

Ba-Ngu Vo, and Ba-Tuong Vo

***Abstract*—State space models in which the system state is a finite set–called the multi-object state–have generated considerable interest in recent years. Smoothing for state space models provides better estimation performance than filtering. In multi-object state estimation, the multi-object filtering density can be efficiently propagated forward in time using an analytic recursion known as the Generalized Labeled Multi-Bernoulli (GLMB) recursion. In this work, we introduce a multi-scan version of the GLMB model to accommodate the multi-object posterior recursion, and develop efficient numerical algorithms for computing this so-called multi-scan GLMB posterior.**

***Index Terms*—State estimation, Filtering, Smoothing, Random finite sets, Multi-dimensional assignment, Gibbs sampling**

I. Introduction

In Bayesian estimation for state-space models, smoothing yields significantly better estimates than filtering by using the history of the states rather than the most recent state [1], [2], [3]. Conditional on the observation history, filtering only considers the current state via the *filtering density*, whereas smoothing considers the entire history of the states up to the current time via the *posterior density*. Numerical computation of the filtering and posterior densities have a long history and is still an active area of research [3], [4], [5], [6].

A generalization of state-space models that has attracted substantial interest in recent years is Mahler's Finite Set Statistics (FISST) framework for multi-object system [7], [8], [9], [10]. Instead of a vector, the state of a multi-object system–the *multi-object state*–is a finite set. Numerically, *multi-object state estimation* [9], [10] is far more complex than traditional state estimation due to additional challenges such as false measurements, misdetection and data association uncertainty. Several tractable multi-object filters have been developed, including the Probability Hypothesis Density (PHD) [7], Cardinalized PHD [8], multi-Bernoulli [9], [11], Dirac delta mixture random finite set [12], hybrid Poisson multi-Bernoulli [13], [14], and second-order PHD [15] filters. These filters, however, are not formulated for estimating (multiple) trajectories. Using labels (or identities), the history of the multi-object states, or the *multi-object trajectory*, is equivalent to the set of object trajectories [16], [17]. Consequently, multi-object trajectory estimation can be achieved via estimation (including filtering) of labeled multi-object states.

In multi-object trajectory estimation, the labeled multi-object filtering recursion admits an analytic solution via the Generalized Labeled Multi-Bernoulli (GLMB) model [16], [18], which can be efficiently approximated [19]. Recent research on parametric approximations and extensions [20], [21], [22], [23], [24], as well as applications such as track-before-detect [21], [25] multi-object sensor control [26], [27], [28], simultaneous localization and mapping [29], [30], multi-object data fusion [31], [32], [33], [34], [35], etc., demonstrate the versatility of the GLMB model, and suggest that it is an important tool for multi-object systems.

Since the filtering density only considers information on the current multi-object state, earlier estimates cannot be updated with current data. Consequently, apart from poorer performance compared to smoothing, an important drawback in a multi-target tracking context is *track fragmentation*, where terminated trajectories are picked up again as new evidence from the data emerges. In contrast, the posterior captures all information on the multi-object trajectory and eliminates track fragmentation as well as improving earlier estimates.

Similar to its single-object counterpart, computing the multi-object posterior is a problem of fundamental importance. Even in the single-object case, computing the posterior and its marginals, or smoothing densities, is still an active area of research [36], [37], [38], [39], [40]. The multi-object posterior density is essential for characterizing other variables/parameters pertaining to the underlying set of objects, other than their trajectories. For example, in cell biology experiments, variables such as cell lifetime, birth rate, death rate, migration pattern, say after a drug is administered, are more useful to biologists than cell tracks [41], [42], [43]. The multi-object posterior enables complete statistical characterization of these variables (e.g. in terms of their distributions, moments), whereas the estimated trajectories cannot.

In this paper, we present a multi-scan version of the GLMB model to accommodate the (labeled) multi-object posterior, and develop an efficient numerical algorithm for multi-object smoothing. Interestingly, the (multi-scan) GLMB posterior recursion takes on an even simpler and more intuitive form than the GLMB filtering recursion. In implementation, however, the GLMB posterior recursion is far more challenging. Preliminary results on the multi-scan GLMB recursion have been reported in [44]. The current work provides a comprehensive treatment, and more importantly algorithms for computing the multi-scan GLMB posterior.

Like the GLMB recursion, the multi-scan GLMB recursion needs to be truncated, and as shown in this article, truncation by retaining components with highest weights minimizes the L_1 truncation error. Unlike the GLMB filtering density, finding the significant components of a GLMB posterior is an NP-hard

Acknowledgement: This work is supported by the Australian Research Council under Discovery Projects DP160104662 and DP170104854.

The authors are with the Department of Electrical and Computer Engineering, Curtin University, Bentley, WA 6102, Australia (email: {ba-ngu.vo, ba-tuong.vo}@curtin.edu.au).

multi-dimensional assignment problem. To solve this problem, we extend the Gibbs sampler for the 2-D assignment problem in [19] to higher dimensions. The resulting technique, capable of solving large-scale multi-dimensional assignment problems, can be applied to compute the GLMB posterior off-line in one batch, or recursively as new observations arrive, thereby performing *smoothing-while-filtering* [2].

The remainder of this article is divided into 5 sections. Section II summarizes relevant concepts in Bayesian multi-object state estimation and the GLMB filter. Section III introduces the multi-scan GLMB model and recursion. Section IV presents an implementation of the multi-scan GLMB recursion using Gibbs sampling. Numerical studies are presented in Section V and conclusions are given in Section VI.

II. Background

Following the convention in [16], the list of variables $X_m, X_{m+1}, ..., X_n$ is abbreviated as $X_{m:n}$, and the inner product $\int f(x)g(x)dx$ is denoted by $\langle f, g\rangle$. For a given set S, $1_S(\cdot)$ denotes the indicator function of S, and $\mathcal{F}(S)$ denotes the class of finite subsets of S. For a finite set X, its cardinality (or number of elements) is denoted by $|X|$, and the product $\prod_{x\in X} f(x)$, for some function f, is denoted by the multi-object exponential f^X, with $f^{\emptyset} = 1$. In addition we use

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

for the generalized Kroneker-δ that takes arbitrary arguments.

A. Multi-object States and Trajectories

This subsection summarizes the representation of trajectories via labeled multi-object states.

At time[1] k, an existing object is described by a vector $x \in \mathbb{X}$ and a unique label $\ell = (s, \alpha)$, where s is the *time of birth*, and α is a unique index to distinguish objects born at the same time (see Figure 1). Let $\mathbb{B}_s$ denote the label space for objects born at time s, then the label space for all objects up to time k (including those born prior to k) is the disjoint union $\mathbb{L}_k = \biguplus_{s=0}^{k} \mathbb{B}_s$ (note that $\mathbb{L}_k = \mathbb{L}_{k-1} \uplus \mathbb{B}_k$). Hence, a *labeled state* $\mathbf{x} = (x, \ell)$ at time k is an element of $\mathbb{X}\times\mathbb{L}_k$.

A *trajectory* is a sequence of labeled states with a common label, at consecutive times [16], i.e. a trajectory with label $\ell = (s, \alpha)$ and states $x_s, x_{s+1}, ..., x_t \in \mathbb{X}$, is the sequence

$$\boldsymbol{\tau} = [(x_s, \ell), (x_{s+1}, \ell), ..., (x_t, \ell)]. \tag{1}$$

A *multi-object state* X is a finite subset of $\mathbb{X}$, and a *labeled multi-object* state at time i is a finite subset $\mathbf{X}$ of $\mathbb{X}\times\mathbb{L}_i$ constructed by augmenting the elements of X with *distinct labels*. Specifically, let $\mathcal{L}: \mathbb{X}\times\mathbb{L}_i \to \mathbb{L}_i$ be the projection defined by $\mathcal{L}((x, \ell)) = \ell$, then $\mathbf{X}$ has distinct labels if and only if the *distinct label indicator* $\Delta(\mathbf{X}) \triangleq \delta_{|\mathbf{X}|}[|\mathcal{L}(\mathbf{X})|]$ equals one.

The labeled states, at time i, of a set S of trajectories defined to have distinct labels and kinematic states at each time, form the labeled multi-object state $\mathbf{X}_i = \{\boldsymbol{\tau}(i) : \boldsymbol{\tau} \in S\}$, where $\boldsymbol{\tau}(i)$ denotes the labeled state of trajectory $\boldsymbol{\tau}$ at time i.

[1]This work considers discrete time indices rather than actual times

Consider a sequence $\mathbf{X}_{j:k}$ of such labeled multi-object states in the interval $\{j:k\}$. Let $\mathbf{x}_i^{(\ell)} = (x_i^{(\ell)}, \ell)$ denote the element of $\mathbf{X}_i$ with label $\ell \in \mathcal{L}(\mathbf{X}_i)$, and unlabeled state $x_i^{(\ell)}$. Then the trajectory with label $\ell \in \cup_{i=j}^{k} \mathcal{L}(\mathbf{X}_i)$ is the sequence of states with label ℓ:

$$\mathbf{x}_{s(\ell):t(\ell)}^{(\ell)} = [(x_{s(\ell)}^{(\ell)}, \ell), ..., (x_{t(\ell)}^{(\ell)}, \ell)], \tag{2}$$

where

$$s(\ell) \triangleq \max\{j, \ell[1, 0]^T\} \tag{3}$$

is the start time, in the interval $\{j:k\}$, of label ℓ, and

$$t(\ell) \triangleq s(\ell) + \textstyle\sum_{i=s(\ell)+1}^{k} 1_{\mathcal{L}(\mathbf{X}_i)}(\ell) \tag{4}$$

is the latest time in $\{s(\ell):k\}$ such that label ℓ still exists. Note that $s(\ell)$ and $t(\ell)$ are also functions of j and k. The multi-object state sequence $\mathbf{X}_{j:k}$ can thus be equivalently represented by the set of all such trajectories, i.e.

$$\mathbf{X}_{j:k} \equiv \left\{\mathbf{x}_{s(\ell):t(\ell)}^{(\ell)} : \ell \in \textstyle\bigcup_{i=j}^{k} \mathcal{L}(\mathbf{X}_i)\right\}. \tag{5}$$

The left and right hand sides of (5) are simply different groupings of the labeled states on the interval $\{j:k\}$. The multi-object state sequence groups the labeled states according to time while the set of trajectories groups according to labels (e.g. see Figure 1). Since the (unlabeled) multi-object state is a finite subset of $\mathbb{X}$, no two trajectories in $\mathbf{X}_{j:k}$ share the same kinematic state at any time.

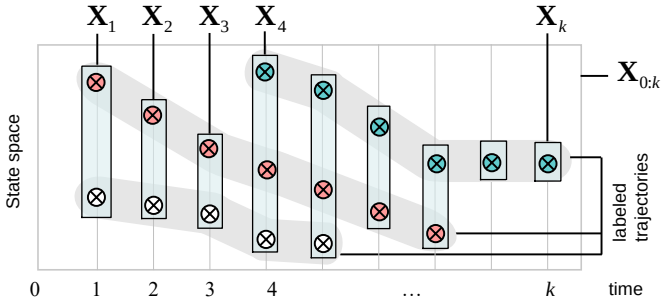

Fig. 1. An example of label assignments (adapted from Figure 1 of [18]). The two objects born at time 1 (red and white) are given labels (1,1) and (1,2), while the only object born at time 4 (blue) is given label (4,1). Note also that the multi-object history $\mathbf{X}_{0:k}$ can be represented by two equivalent groupings: (a) according to time (the vertical strips containing states of the same time, i.e. the multi-object states) or; (b) according to labels (grey strips containing states of the same color, i.e. the trajectories).

Hereon, single-object states are represented by lowercase letters (e.g. x, $\mathbf{x}$), multi-object states are represented by uppercase letters (e.g. X, $\mathbf{X}$), symbols for labeled states and their distributions are bolded to distinguish them from unlabeled ones (e.g. $\mathbf{x}$, $\mathbf{X}$, $\boldsymbol{\pi}$, etc.). The term multi-object state refers to both unlabeled or label multi-object states (the context is clear from the bolded or unbolded symbols). Given a sequence of sets $I_{j:k}$, with a slight abuse of notation, the union $\cup_{i=j}^{k} I_i$ is written as $I_{j:k}$. The context should be clear when we write $I \subseteq I_{j:k}$, $\ell \in I_{j:k}$, and $\mathcal{F}(I_{j:k})$. Similarly, we use $\mathcal{L}(\mathbf{X}_{j:k}) \triangleq (\mathcal{L}(\mathbf{X}_j), ..., \mathcal{L}(\mathbf{X}_k))$ in place of $\cup_{i=j}^{k} \mathcal{L}(\mathbf{X}_i)$.

B. Bayes Recursion

Following the Bayesian paradigm, each labeled multi-object state is modeled as a labeled random finite set (RFS) [16], characterized by the Finite Set Statistics (FISST) multi-object density [7], [45]. A labeled RFS is defined as a marked RFS with distinct marks [16], i.e., a labeled RFS with state space $\mathbb{X}$ and label space $\mathbb{L}$ is an RFS of $\mathbb{X}\times\mathbb{L}$, constructed by marking the elements of an RFS of $\mathbb{X}$ with distinct labels from $\mathbb{L}$ (note that the 2nd clause is implicit from Definition 1 of [16]).

Given the observation history $Z_{1:k}$, all information on the set of objects is captured in the *multi-object posterior density* $\boldsymbol{\pi}_{0:k}(\mathbf{X}_{0:k}) \triangleq \boldsymbol{\pi}_{0:k}(\mathbf{X}_{0:k}|Z_{1:k})$. Note that the dependence on $Z_{1:k}$ is omitted for notational compactness. Similar to standard Bayesian state estimation [2], [3], the (multi-object) posterior density can be propagated forward recursively by

$$\boldsymbol{\pi}_{0:k}(\mathbf{X}_{0:k}) = \frac{g_k(Z_k|\mathbf{X}_k)\mathbf{f}_{k|k-1}(\mathbf{X}_k|\mathbf{X}_{k-1})\boldsymbol{\pi}_{0:k-1}(\mathbf{X}_{0:k-1})}{h_k(Z_k|Z_{1:k-1})}, \tag{6}$$

where $g_k(\cdot|\cdot)$ is the *multi-object likelihood function* at time k, $\mathbf{f}_{k|k-1}(\cdot|\cdot)$ is the *multi-object transition density* to time k, and $h_k(Z_k|Z_{1:k-1})$ is the normalizing constant or predictive likelihood. A well-defined $\mathbf{f}_{k|k-1}(\cdot|\cdot)$ ensures that the multi-object history $\mathbf{X}_{0:k}$ represents a set of trajectories [16].

Markov Chain Monte Carlo (MCMC) approximations of the multi-object posterior have been proposed in [17] for detection measurements using Particle MCMC [46], and in [47] for image measurements using reversible jump MCMC.

A cheaper alternative is the *multi-object filtering density*, $\boldsymbol{\pi}_k(\mathbf{X}_k) \triangleq \int \boldsymbol{\pi}_{0:k}(\mathbf{X}_{0:k})\delta\mathbf{X}_{0:k-1}$, which can be propagated by the *multi-object Bayes filter* [7], [9]

$$\boldsymbol{\pi}_k(\mathbf{X}_k) = \frac{g_k(Z_k|\mathbf{X}_k)\int \mathbf{f}_{k|k-1}(\mathbf{X}_k|\mathbf{X}_{k-1})\boldsymbol{\pi}_{k-1}(\mathbf{X}_{k-1})\delta\mathbf{X}_{k-1}}{h_k(Z_k|Z_{1:k-1})}.$$

The GLMB filter is an analytic solution to this recursion under the standard multi-object system model [16]. For more general models, the multi-object particle filter [25] can be used.

C. Multi-Object System Model

Given a multi-object state $\mathbf{X}_{k-1}$ (at time $k-1$), each state $\mathbf{x}_{k-1} = (x_{k-1}, \ell_{k-1}) \in \mathbf{X}_{k-1}$ either survives with probability $P_{S,k-1}(\mathbf{x}_{k-1})$ and evolves to a new state (x_k, ℓ_k) with probability density $f_{S,k|k-1}(x_k|x_{k-1}, \ell_{k-1})\delta_{\ell_{k-1}}[\ell_k]$ or dies with probability $Q_{S,k-1}(\mathbf{x}_{k-1}) = 1 - P_{S,k-1}(\mathbf{x}_{k-1})$. Further, for each ℓ_k in a (finite) birth label space $\mathbb{B}_k$ at time k, either a new object with state (x_k, ℓ_k) is born with probability $P_{B,k}(\ell_k)$ and density $f_{B,k}(x_k, \ell_k)$, or unborn with probability $Q_{B,k}(\ell_k) = 1-P_{B,k}(\ell_k)$. The multi-object state $\mathbf{X}_k$ (at time k) is the superposition of surviving states and new born states. The multi-object transition density $\mathbf{f}_{k|k-1}(\mathbf{X}_k|\mathbf{X}_{k-1})$, given by equation (6) of [18], is constructed from Mahler's multi-object transition density by marking (unlabeled) multi-object states with distinct labels. Hence, no two elements of a multi-object state shares the same kinematic or unlabeled state.

Given a multi-object state $\mathbf{X}_k$, each $\mathbf{x}_k \in \mathbf{X}_k$ is either detected with probability $P_{D,k}(\mathbf{x}_k)$ and generates a detection z with likelihood $g_{D,k}(z|\mathbf{x}_k)$ or missed with probability $Q_{D,k}(\mathbf{x}_k) = 1-P_{D,k}(\mathbf{x}_k)$. The *multi-object observation* Z_k is the superposition of the observations from detected objects and Poisson clutter with intensity κ_k. Assuming that, conditional on $\mathbf{X}_k$, detections are independent of each other and clutter, the multi-object likelihood function is given by [16], [18]

$$g_k(Z_k|\mathbf{X}_k) \propto \sum_{\theta_k\in\Theta_k} 1_{\Theta_k(\mathcal{L}(\mathbf{X}_k))}(\theta_k)\left[\psi_{k,Z_k}^{(\theta_k\circ\mathcal{L}(\cdot))}(\cdot)\right]^{\mathbf{X}_k} \tag{7}$$

where Θ_k denotes the set of maps $\theta_k : \mathbb{L}_k \to \{0{:}|Z_k|\}$ that are *positive 1-1* (i.e. θ_k never *map distinct arguments to the same positive value*), $\Theta_k(I)$ denotes the subset of Θ_k with domain I, $\theta_k \circ \mathcal{L}(\mathbf{x}) = \theta_k(\mathcal{L}(\mathbf{x}))$, and

$$\psi_{k,\{z_{1:m}\}}^{(i)}(x,\ell) = \begin{cases} \frac{P_{D,k}(x,\ell)g_{D,k}(z_i|x,\ell)}{\kappa_k(z_i)}, & i>0 \\ Q_{D,k}(x,\ell), & i=0 \end{cases}, \tag{8}$$

The map θ_k assigns a detected label ℓ to measurement $z_{\theta_k(\ell)} \in Z_k$, while for an undetected label $\theta_k(\ell) = 0$.

D. GLMB Filtering Recursion

A *generalized labeled multi-Bernoulli* (GLMB) density on $\mathcal{F}(\mathbb{X}\times\mathbb{L})$ has the form [16]:

$$\boldsymbol{\pi}(\mathbf{X}) = \Delta(\mathbf{X})\sum_{\xi\in\Xi} w^{(\xi)}(\mathcal{L}(\mathbf{X}))[p^{(\xi)}]^{\mathbf{X}}, \tag{9}$$

where Ξ is a discrete index set, each $p^{(\xi)}(\cdot,\ell)$ is a probability density on $\mathbb{X}$, i.e., $\int p^{(\xi)}(x,\ell)dx = 1$, and each $w^{(\xi)}(L)$ is non-negative with $\sum_{L\subseteq\mathbb{L}}\sum_{\xi\in\Xi} w^{(\xi)}(L) = 1$. The GLMB density (9) can be interpreted as a mixture of multi-object exponentials, where the weights are functions of the labels.

The GLMB family is closed under the Bayes multi-object filtering recursion and an explicit expression relating the filtering density at time k to that at time $k-1$ is given by (14) of [19]. This recursion can be expressed as follows.

Given the GLMB filtering density at time $k-1$,

$$\boldsymbol{\pi}_{k-1}(\mathbf{X}_{k-1}) = \Delta(\mathbf{X}_{k-1})\sum_{\xi\in\Xi} w_{k-1}^{(\xi)}(\mathcal{L}(\mathbf{X}_{k-1}))[p_{k-1}^{(\xi)}]^{\mathbf{X}_{k-1}}, \tag{10}$$

the GLMB filtering density at time k is

$$\begin{aligned}\boldsymbol{\pi}_k(\mathbf{X}_k) &\propto \Delta(\mathbf{X}_k)\sum_{\xi,\theta_k,I_{k-1}} \omega_k^{(\xi,\theta_k)}(I_{k-1})\delta_{\mathcal{D}(\theta_k)}[\mathcal{L}(\mathbf{X}_k)][p_k^{(\xi,\theta_k)}]^{\mathbf{X}_k}, \\ \omega_k^{(\xi,\theta_k)}(I_{k-1}) &= 1_{\mathcal{F}(\mathbb{B}_k\uplus I_{k-1})}(\mathcal{D}(\theta_k))[\omega_{k|k-1}^{(\xi,\theta_k)}]^{\mathbb{B}_k\uplus I_{k-1}} w_{k-1}^{(\xi)}(I_{k-1})\end{aligned} \tag{11, 12}$$

$$\omega_{k|k-1}^{(\xi,\theta_k)}(\ell) = \begin{cases} \bar{\Lambda}_{B,k}^{(\theta_k(\ell))}(\ell), & \ell\in\mathcal{D}(\theta_k)\cap\mathbb{B}_k \\ \bar{\Lambda}_{S,k|k-1}^{(\xi,\theta_k(\ell))}(\ell), & \ell\in\mathcal{D}(\theta_k)-\mathbb{B}_k \\ Q_{B,k}(\ell), & \ell\in\overline{\mathcal{D}(\theta_k)}\cap\mathbb{B}_k \\ \bar{Q}_{S,k-1}^{(\xi)}(\ell), & \ell\in\overline{\mathcal{D}(\theta_k)}-\mathbb{B}_k \end{cases}, \tag{13}$$

$$p_k^{(\xi,\theta_k)}(x,\ell) = \begin{cases} \frac{\Lambda_{B,k}^{(\theta_k(\ell))}(x,\ell)}{\bar{\Lambda}_{B,k}^{(\theta_k(\ell))}(\ell)}, & \ell\in\mathcal{D}(\theta_k)\cap\mathbb{B}_k \\ \frac{\left\langle\Lambda_{S|k-1}^{(\theta_k(\ell))}(x|\cdot,\ell), p_{k-1}^{(\xi)}(\cdot,\ell)\right\rangle}{\bar{\Lambda}_{S,k|k-1}^{(\xi,\theta_k(\ell))}(\ell)}, & \ell\in\mathcal{D}(\theta_k)-\mathbb{B}_k \end{cases}, \tag{14}$$

$$\Lambda^{(j)}_{B,k}(x,\ell) = \psi^{(j)}_{k,Z_k}(x,\ell) f_{B,k}(x,\ell) P_{B,k}(\ell), \quad (15)$$

$$\Lambda^{(j)}_{S,k|k-1}(x|\varsigma,\ell) = \psi^{(j)}_{k,Z_k}(x,\ell) f_{S,k|k-1}(x|\varsigma,\ell) P_{S,k-1}(\varsigma,\ell), \quad (16)$$

$$\bar{Q}^{(\xi)}_{S,k-1}(\ell) = \langle Q_{S,k-1}(\cdot,\ell), p^{(\xi)}_{k-1}(\cdot,\ell)\rangle, \quad (17)$$

$$\bar{\Lambda}^{(j)}_{B,k}(\ell) = \langle \Lambda^{(j)}_{B,k}(\cdot,\ell), 1\rangle, \quad (18)$$

$$\bar{\Lambda}^{(\xi,j)}_{S,k|k-1}(\ell) = \int \langle \Lambda^{(j)}_{S,k|k-1}(x|\cdot,\ell), p^{(\xi)}_{k-1}(\cdot,\ell)\rangle dx. \quad (19)$$

III. GLMB Posterior Recursion

This section presents a multi-scan version of the GLMB model, and subsequently a multi-scan GLMB recursion.

A. Multi-Scan GLMB

Recall the equivalence between the multi-object state sequence $\mathbf{X}_{j:k}$ and the set $\{\mathbf{x}^{(\ell)}_{s(\ell):t(\ell)} : \ell \in \mathcal{L}(\mathbf{X}_{j:k})\}$ of trajectories in (5). Noting from (2) that $\mathbf{x}^{(\ell)}_{s(\ell):t(\ell)}$ is completely characterized by ℓ and its kinematic states $x^{(\ell)}_{s(\ell):t(\ell)}$, we use

$$\mathbf{x}^{(\ell)}_{s(\ell):t(\ell)} \equiv (x^{(\ell)}_{s(\ell):t(\ell)}, \ell).$$

For any function h taking the trajectories to the non-negative reals, we introduce the following so-called *multi-scan exponential* notation:

$$[h]^{\mathbf{X}_{j:k}} \triangleq [h]^{\left\{\mathbf{x}^{(\ell)}_{s(\ell):t(\ell)}:\, \ell\in\mathcal{L}(\mathbf{X}_{j:k})\right\}} = \prod_{\ell\in\mathcal{L}(\mathbf{X}_{j:k})} h(\mathbf{x}^{(\ell)}_{s(\ell):t(\ell)}) \quad (20)$$

This notation is quite suggestive of the exponential property in the following Lemma (see Appendix VII-B for the proof).

Lemma 1. *Let $\mathbf{X}_{j:k}$ be a sequence of multi-object states (generated by a set of trajectories) and g, h be two functions taking trajectories to the reals. Then:*

(i) $[g\ h]^{\mathbf{X}_{j:k}} = [g]^{\mathbf{X}_{j:k}}\,[h]^{\mathbf{X}_{j:k}}$

(ii) For a multi-object state sequence $\mathbf{Y}_{j:k}$ with labels disjoint from those of $\mathbf{X}_{j:k}$,

$$[h]^{\mathbf{X}_{j:k}\uplus\mathbf{Y}_{j:k}} = [h]^{\mathbf{X}_{j:k}}\,[h]^{\mathbf{Y}_{j:k}}$$

(iii) For any i in $\{j{:}k\}$

$$[g]^{\mathbf{X}_{j:i}}\,[h]^{\mathbf{X}_{i:k}} = [g\odot h]^{\mathbf{X}_{j:k}},$$

where

$$(g\odot h)(\mathbf{x}^{(\ell)}_{s(\ell):t(\ell)}) = \begin{cases} h(\mathbf{x}^{(\ell)}_{s(\ell):t(\ell)}), & i < s(\ell) \\ g(\mathbf{x}^{(\ell)}_{s(\ell):i}) h(\mathbf{x}^{(\ell)}_{i:t(\ell)}), & s(\ell)\le i\le t(\ell) \\ g(\mathbf{x}^{(\ell)}_{s(\ell):t(\ell)}), & t(\ell) < i \end{cases},$$

$s(\ell)$ and $t(\ell)$, given by (3) and (4) are, respectively, the starting and terminating times on $\{j:k\}$ for label ℓ.

Using multi-scan exponential notation, the multi-object transition density given in [16] can be written tersely as follows (for completeness the proof is given in Appendix VII-C).

Proposition 2. *For the multi-object dynamic model described in subsection II-C, the multi-object transition density is*

$$\mathbf{f}_{k|k-1}(\mathbf{X}_k|\mathbf{X}_{k-1}) = \quad (21)$$
$$\Delta(\mathbf{X}_k) 1_{\mathcal{F}(\mathbb{B}_k\uplus\mathcal{L}(\mathbf{X}_{k-1}))}(\mathcal{L}(\mathbf{X}_k)) Q^{\mathbb{B}_k-\mathcal{L}(\mathbf{X}_k)}_{B,k} \left[\phi_{k-1:k}\right]^{\mathbf{X}_{k-1:k}}$$

where

$$\phi_{k-1:k}(x^{(\ell)}_{s(\ell):t(\ell)}, \ell) = \quad (22)$$
$$\begin{cases} P_{B,k}(\ell) f_{B,k}(x^{(\ell)}_k, \ell), & s(\ell) = k \\ P_{S,k-1}(x^{(\ell)}_{k-1}, \ell) f_{S,k|k-1}(x^{(\ell)}_k | x^{(\ell)}_{k-1}, \ell), & t(\ell) = k > s(\ell) \\ Q_{S,k-1}(x^{(\ell)}_{k-1}, \ell), & t(\ell) = k-1 \end{cases}$$

Definition 1. A density on $\mathcal{F}(\mathbb{X}\times\mathbb{L}_j)\times ...\times\mathcal{F}(\mathbb{X}\times\mathbb{L}_k)$ is *multi-scan GLMB* if it has the form

$$\boldsymbol{\pi}(\mathbf{X}_{j:k}) = \Delta(\mathbf{X}_{j:k}) \sum_{\xi\in\Xi} w^{(\xi)}_{j:k}(\mathcal{L}(\mathbf{X}_{j:k}))[p^{(\xi)}_{j:k}]^{\mathbf{X}_{j:k}} \quad (23)$$

where: Ξ is a discrete set; $\Delta(\mathbf{X}_{j:k}) \triangleq \prod_{i=j}^{k}\Delta(\mathbf{X}_i)$; $w^{(\xi)}_{j:k}(I_{j:k})$, $(\xi, I_{j:k}) \in \Xi\times\mathcal{F}(\mathbb{L}_j)\times ...\times\mathcal{F}(\mathbb{L}_k)$ is non-negative with

$$\sum_{\xi, I_{j:k}} w^{(\xi)}_{j:k}(I_{j:k}) = 1; \quad (24)$$

and $p^{(\xi)}_{j:k}(\cdot,\ell)$, $\xi\in\Xi$, $\ell\in\mathcal{L}(\mathbf{X}_{j:k})$, defined on $\mathbb{X}^{t(\ell)-s(\ell)+1}$ with starting and terminating times $s(\ell)$ and $t(\ell)$ on $\{j:k\}$, given by (3) and (4), satisfies

$$\int p^{(\xi)}_{j:k}(x_{s(\ell):t(\ell)}, \ell) dx_{s(\ell):t(\ell)} = 1. \quad (25)$$

Similar to the GLMB, the multi-scan GLMB (23) can be expressed in the so-called δ-form:

$$\boldsymbol{\pi}(\mathbf{X}_{j:k}) = \Delta(\mathbf{X}_{j:k}) \sum_{\xi, I_{j:k}} w^{(\xi)}_{j:k}(I_{j:k}) \delta_{I_{j:k}}[\mathcal{L}(\mathbf{X}_{j:k})][p^{(\xi)}_{j:k}]^{\mathbf{X}_{j:k}} \quad (26)$$

Each term or component of a multi-scan GLMB consists of a weight $w^{(\xi)}_{j:k}(I_{j:k})$ and a multi-scan exponential $[p^{(\xi)}_{j:k}]^{\mathbf{X}_{j:k}}$ with label history that matches $I_{j:k}$. The weight $w^{(\xi)}_{j:k}(I_{j:k})$ can be interpreted as the probability of hypothesis $(\xi, I_{j:k})$, and for each $\ell\in I_{j:k}$, $p^{(\xi)}_{j:k}(\cdot,\ell)$ is the joint probability density of its (trajectory) kinematic states, given hypothesis $(\xi, I_{j:k})$.

Proposition 3. *For a function $f : \mathcal{F}(\mathbb{L}_j)\times ...\times\mathcal{F}(\mathbb{L}_k)\to\mathbb{R}$, its integral with respect to the multi-scan GLMB (23) is*

$$\int f(\mathcal{L}(\mathbf{X}_{j:k}))\boldsymbol{\pi}(\mathbf{X}_{j:k})\delta\mathbf{X}_{j:k} = \sum_{\xi, I_{j:k}} f(I_{j:k}) w^{(\xi)}_{j:k}(I_{j:k}) \quad (27)$$

where $(\xi, I_{j:k}) \in \Xi\times\mathcal{F}(\mathbb{L}_j)\times ...\times\mathcal{F}(\mathbb{L}_k)$. See Appendix VII-D for proof.

By setting f to 1 in the above proposition, the multi-scan GLMB integrates to 1, and hence, is a FISST density. Some useful statistics from the multi-scan GLMB follow from the above proposition for suitably defined functions of the labels.

Corollary 4. *The cardinality distribution, i.e. distribution of the number of trajectories is given by*

$$\Pr(|\mathcal{L}(\mathbf{X}_{j:k})| = n) = \sum_{\xi, I_{j:k}} \delta_n[|I_{j:k}|] w^{(\xi)}_{j:k}(I_{j:k}) \quad (28)$$

Corollary 5. *The joint probability of existence of a non-empty set of trajectories with labels L is given by*

$$\Pr(L \textit{ exist}) = \sum_{\xi, I_{j:k}} 1_{\mathcal{F}(I_{j:k})}(L) w^{(\xi)}_{j:k}(I_{j:k}), \quad (29)$$

and as a special case

$$\Pr(\ell \textit{ exists}) = \sum_{\xi, I_{j:k}} 1_{I_{j:k}}(\ell) w^{(\xi)}_{j:k}(I_{j:k}). \tag{30}$$

Corollary 6. *The distribution of trajectory lengths is given by*

$$\Pr(\textit{a trajectory has length } m) = \sum_{\xi, I_{j:k}} \frac{w^{(\xi)}_{j:k}(I_{j:k})}{|I_{j:k}|} \sum_{\ell \in I_{j:k}} \delta_m[t(\ell) - s(\ell) + 1], \tag{31}$$

and the distribution of the length of trajectory with label ℓ is

$$\Pr(\textit{length}(\ell) = m) = \sum_{\xi, I_{j:k}} \delta_m[t(\ell) - s(\ell) + 1]\, 1_{I_{j:k}}(\ell) w^{(\xi)}_{j:k}(I_{j:k}). \tag{32}$$

Corollary 7. *The cardinality distributions of births and deaths at time $u \in \{j:k\}$ are given by*

$$\Pr(n \textit{ births at time } u) = \sum_{\xi, I_{j:k}} w^{(\xi)}_{j:k}(I_{j:k}) \delta_n\left[\textstyle\sum_{\ell \in I_{j:k}} \delta_u(s(\ell))\right], \tag{33}$$

$$\Pr(n \textit{ deaths at time } u) = \sum_{\xi, I_{j:k}} w^{(\xi)}_{j:k}(I_{j:k}) \delta_n\left[\textstyle\sum_{\ell \in I_{j:k}} \delta_u(t(\ell))\right], \tag{34}$$

Similar to its single-scan counterpart, a number of estimators can be constructed for a multi-scan GLMB. The simplest would be to find the multi-scan GLMB component with the highest weight $w^{(\xi)}_{j:k}(I_{j:k})$ and compute the most probable or expected trajectory estimate from $p^{(\xi)}_{j:k}(\cdot, \ell)$ for each $\ell \in I_{j:k}$. Alternatively, instead of the most significant, we can use the most significant amongst components with the most probable cardinality n^* (determined by maximizing (28)).

Another class of estimators, based on existence probabilities, can be constructed as follows. Find the set of labels L^* with highest joint existence probability by maximizing (29). Alternatively, we can choose L^* as the label set of cardinality n^* with highest joint existence probability, or the set of n^* labels with highest individual existence probabilities. For each $\ell \in L^*$, we determine the most probable length m^* by maximizing (32) and compute the trajectory density

$$p_{j:k}(\cdot, \ell) \propto \sum_{\xi, I_{j:k}} \delta_{m^*}[t(\ell) - s(\ell) + 1]\, 1_{I_{j:k}}(\ell) w^{(\xi)}_{j:k}(I_{j:k}) p^{(\xi)}_{j:k}(\cdot, \ell),$$

from which the mode or mean trajectory can be determined.

B. Closure Under Bayes Recursion

Conceptually, a multi-scan GLMB is simply a GLMB where the argument is a set of labeled trajectories represented by $\mathbf{X}_{j:k}$, and hence is closed under the Bayes recursion [16].

Multiplying the multi-scan GLMB

$$\boldsymbol{\pi}(\mathbf{X}_{j:k-1}) = \Delta(\mathbf{X}_{j:k-1}) \sum_{\xi \in \Xi} w^{(\xi)}_{j:k-1}(\mathcal{L}(\mathbf{X}_{j:k-1})) \left[p^{(\xi)}_{j:k-1}\right]^{\mathbf{X}_{j:k-1}}, \tag{35}$$

by $\mathbf{f}_{k|k-1}(\mathbf{X}_k|\mathbf{X}_{k-1})$ in (21), and using Lemma 1 (iii) to "stitch" $[p^{(\xi)}_{j:k-1}]^{\mathbf{X}_{j:k-1}}$ and $\left[\phi_{k-1:k}\right]^{\mathbf{X}_{k-1:k}}$ together, yields a multi-scan GLMB prediction density of the form (23), with

$$p^{(\xi)}_{j:k} = p^{(\xi)}_{j:k-1} \odot \phi_{k-1:k}, \tag{36}$$

$$w^{(\xi)}_{j:k}(I_{j:k}) = w^{(\xi)}_{j:k-1}(I_{j:k-1}) 1_{\mathcal{F}(\mathbb{B}_k \uplus I_{k-1})}(I_k) Q^{\mathbb{B}_k - I_k}_{B,k}. \tag{37}$$

Further, the measurement likelihood for $\mathbf{X}_{j:k}$ has the same exponential-mixture form as (54) in [16], i.e.

$$g_k(Z_k|\mathbf{X}_{j:k}) \propto \sum_{\theta_k \in \Theta_k} 1_{\Theta_k(\mathcal{L}(\mathbf{X}_k))}(\theta_k) \left[\psi^{(\theta_k \circ \mathcal{L}(\cdot))}_{j:k, Z_k}(\cdot)\right]^{\mathbf{X}_{j:k}},$$

with

$$\psi^{(i)}_{j:k, Z_k}(x_{s(\ell):t(\ell)}, \ell) = \begin{cases} \psi^{(i)}_{k, Z_k}(x_{t(\ell)}, \ell), & t(\ell) = k \\ 1, & t(\ell) < k \end{cases},$$

where $\theta_k \circ \mathcal{L}(\boldsymbol{\tau}) = \theta_k(\mathcal{L}(\boldsymbol{\tau}))$, and $t(\ell)$, given by (4), is the terminating time on $\{j:k\}$ for label ℓ (since trajectories terminated before time k do not contribute to Z_k). Thus, the multi-scan GLMB prior (23) is closed under Bayes update, and its posterior given by (19) of [16], i.e.,

$$\boldsymbol{\pi}(\mathbf{X}_{j:k}|Z_k) \propto \Delta(\mathbf{X}_{j:k}) \sum_{\xi, \theta_k} w^{(\xi, \theta_k)}_{j:k}(\mathcal{L}(\mathbf{X}_{j:k})) \left[p^{(\xi, \theta_k)}_{j:k}\right]^{\mathbf{X}_{j:k}} \tag{38}$$

where $\xi \in \Xi$, $\theta_k \in \Theta_k$,

$$p^{(\xi, \theta_k)}_{j:k}(\cdot, \ell) = \frac{p^{(\xi)}_{j:k}(\cdot, \ell) \psi^{(\theta_k(\ell))}_{j:k, Z_k}(\cdot, \ell)}{\bar{\psi}^{(\xi, \theta_k)}_{j:k, Z_k}(\ell)} \tag{39}$$

$$\bar{\psi}^{(\xi, \theta_k)}_{j:k, Z_k}(\ell) = \left\langle p^{(\xi)}_{j:k}(\cdot, \ell), \psi^{(\theta_k(\ell))}_{j:k, Z_k}(\cdot, \ell) \right\rangle \tag{40}$$

$$w^{(\xi, \theta_k)}_{j:k}(I_{j:k}) = 1_{\Theta_k(I_k)}(\theta_k) w^{(\xi)}_{j:k}(I_{j:k}) \left[\bar{\psi}^{(\xi, \theta_k)}_{j:k, Z_k}\right]^{I_{j:k}} \tag{41}$$

Remark: A similar version of the GLMB Bayes update (38) was derived in [48] using a sophisticated approach based on probability density of random sets of trajectories. Interestingly, [48] is the first to undertake the development of probability densities on the space of all finite subsets of the disjoint union of Cartesian products of $\mathbb{X} \times \mathbb{L}$ (or $\mathbb{X}$). This is fundamentally different from the standard state-space estimation paradigm followed in our work, where densities are defined on the Cartesian product of the state-space [2], [3], [49]. Other main differences from (38) are: sample sets of trajectories do not necessarily have distinct labels; and the weights in the posterior probability density are independent of the argument.

C. Multi-Scan GLMB Posterior Recursion

Setting $j = 0$ in the GLMB update (38)-(41) and expanding the parameters (for completeness see Appendix VII-E) yields an explicit multi-scan GLMB posterior recursion.

Proposition 8. *Under the standard multi-object system model, if the multi-object posterior at time $k-1$ is*

$$\boldsymbol{\pi}_{0:k-1}(\mathbf{X}_{0:k-1}) = \Delta(\mathbf{X}_{0:k-1}) \sum_{\xi \in \Xi} w^{(\xi)}_{0:k-1}(\mathcal{L}(\mathbf{X}_{0:k-1})) [p^{(\xi)}_{0:k-1}]^{\mathbf{X}_{0:k-1}}, \tag{42}$$

then the multi-object posterior at time k is

$$\boldsymbol{\pi}_{0:k}(\mathbf{X}_{0:k}) \propto \Delta(\mathbf{X}_{0:k}) \sum_{\xi,\theta_k} \omega_{0:k}^{(\xi,\theta_k)}(\mathcal{L}(\mathbf{X}_{0:k-1}))\delta_{\mathcal{D}(\theta_k)}[\mathcal{L}(\mathbf{X}_k)][p_{0:k}^{(\xi,\theta_k)}]^{\mathbf{X}_{0:k}} \tag{43}$$

where $\xi \in \Xi$, $\theta_k \in \Theta_k$, $\mathcal{D}(\theta_k)$ is the domain of θ_k,

$$\omega_{0:k}^{(\xi,\theta_k)}(I_{0:k-1}) = 1_{\mathcal{F}(\mathbb{B}_k \uplus I_{k-1})}(\mathcal{D}(\theta_k))\left[\omega_{k|k-1}^{(\xi,\theta_k)}\right]^{\mathbb{B}_k \uplus I_{k-1}} w_{0:k-1}^{(\xi)}(I_{0:k-1}) \tag{44}$$

$$\omega_{k|k-1}^{(\xi,\theta_k)}(\ell) = \begin{cases} \bar{\Lambda}_{B,k}^{(\theta_k(\ell))}(\ell), & \ell \in \mathcal{D}(\theta_k) \cap \mathbb{B}_k \\ \bar{\Lambda}_{S,k|k-1}^{(\xi,\theta_k(\ell))}(\ell), & \ell \in \mathcal{D}(\theta_k) - \mathbb{B}_k \\ Q_{B,k}(\ell), & \ell \in \overline{\mathcal{D}(\theta_k)} \cap \mathbb{B}_k \\ \bar{Q}_{S,k-1}^{(\xi)}(\ell), & \ell \in \overline{\mathcal{D}(\theta_k)} - \mathbb{B}_k \end{cases}, \tag{45}$$

$$p_{0:k}^{(\xi,\theta_k)}(x_{s(\ell):t(\ell)}, \ell) = \begin{cases} \frac{\Lambda_{B,k}^{(\theta_k(\ell))}(x_k,\ell)}{\bar{\Lambda}_{B,k}^{(\theta_k(\ell))}(\ell)}, & s(\ell) = k \\ \frac{\Lambda_{S,k|k-1}^{(\theta_k(\ell))}(x_k|x_{k-1},\ell)p_{0:k-1}^{(\xi)}(x_{s(\ell):k-1},\ell)}{\bar{\Lambda}_{S,k|k-1}^{(\xi,\theta_k(\ell))}(\ell)}, & t(\ell) = k > s(\ell) \\ \frac{Q_{S,k-1}(x_{k-1},\ell)p_{0:k-1}^{(\xi)}(x_{s(\ell):k-1},\ell)}{\bar{Q}_{S,k-1}^{(\xi)}(\ell)}, & t(\ell) = k-1 \\ p_{0:k-1}^{(\xi)}(x_{s(\ell):t(\ell)},\ell), & t(\ell) < k-1 \end{cases} \tag{46}$$

The multi-scan GLMB posterior recursion (42)-(46) bears remarkable resemblance to the GLMB filtering recursion (10)-(15). In essence, it is the GLMB filtering recursion without the marginalization of past labels and kinematic states. Indeed, the weight increments for multi-scan GLMB and GLMB components are identical. Arguably, the multi-scan GLMB recursion is more intuitive because it does not involve marginalization over previous label sets nor past states of the trajectories.

The multi-scan GLMB recursion initiates new born trajectories, updates surviving trajectories, terminates disappearing trajectories, and stores trajectories that disappeared earlier. Noting that $\ell \in \mathcal{D}(\theta_k) \cap \mathbb{B}_k$ is equivalent to $s(\ell) = k$, initiation of new trajectories is identical to that of the GLMB filter. Noting $\ell \in \mathcal{D}(\theta_k) - \mathbb{B}_k$ is equivalent to $t(\ell) = k > s(\ell)$, the update of surviving trajectories is the same as the GLMB filter, but without marginalization of past kinematic states. On the other hand, termination/storing of disappearing/disappeared trajectories are not needed in the GLMB filter.

IV. Computing Multi-Scan GLMB Posteriors

The number of terms of the multi-scan GLMB posterior grows super-exponentially in time and it is necessary to find a tractable approximation. Functional approximation criteria, e.g. L_p-norm, or divergences such as Kullback-Leibler [21], Renyi, Cauchy-Schwarz [50], [26], can be extended to the multi-scan case. The main challenge is: given a prescribed number of terms, what is the best multi-scan GLMB approximation of the posterior, without exhaustive enumeration?

Using the same lines of arguments as Proposition 5 of [18], the L_1-error between a multi-scan GLMB and its truncation is given by the following result.

Proposition 9. *Let $\|\mathbf{f}\|_1 \triangleq \int |\mathbf{f}(\mathbf{X}_{j:k})| \delta\mathbf{X}_{j:k}$ denote the L_1-norm of $\mathbf{f} : \mathcal{F}(\mathbb{X}\times\mathbb{L}_j)\times...\times\mathcal{F}(\mathbb{X}\times\mathbb{L}_k) \to \mathbb{R}$, and for a given $\mathbb{H} \subseteq \Xi\times\mathcal{F}(\mathbb{L}_j)\times...\times\mathcal{F}(\mathbb{L}_k)$ let*

$$\mathbf{f}_{\mathbb{H}}(\mathbf{X}_{j:k}) \triangleq \Delta(\mathbf{X}_{j:k}) \sum_{(\xi,I_{j:k})\in\mathbb{H}} w^{(\xi)}(I_{j:k})\delta_{I_{j:k}}[\mathcal{L}(\mathbf{X}_{j:k})][p^{(\xi)}]^{\mathbf{X}_{j:k}}$$

be an unnormalized multi-scan GLMB density. If $\mathbb{T} \subseteq \mathbb{H}$ then

$$\|\mathbf{f}_{\mathbb{H}} - \mathbf{f}_{\mathbb{T}}\|_1 = \sum_{(\xi,I_{j:k})\in\mathbb{H}-\mathbb{T}} w^{(\xi)}(I_{j:k}),$$

$$\left\| \frac{\mathbf{f}_{\mathbb{H}}}{||\mathbf{f}_{\mathbb{H}}||_1} - \frac{\mathbf{f}_{\mathbb{T}}}{||\mathbf{f}_{\mathbb{T}}||_1} \right\|_1 \leq 2\frac{||\mathbf{f}_{\mathbb{H}}||_1 - ||\mathbf{f}_{\mathbb{T}}||_1}{||\mathbf{f}_{\mathbb{H}}||_1}.$$

Hence, given a multi-scan GLMB posterior, the minimum L_1-norm approximation for a prescribed number of terms can be obtained by keeping only those with highest weights. Furthermore, this can be accomplished, without exhaustive enumeration, by solving the multi-dimensional assignment problem [51]. This problem is NP-hard for more than two scans, and state-of-the-art algorithms cannot handle more than 10 scans with about 20 measurements per scan; see for example [52] and references therein.

This section presents efficient techniques for computing multi-scan GLMB posteriors by Gibbs sampling. Subsection IV-A formulates multi-scan GLMB posterior truncation as a multi-dimensional assignment problem, while solutions are developed in subsections IV-B and IV-C.

A. Canonical Multi-Scan GLMB Posterior

To express the multi-scan GLMB posterior in canonical form, we represent each $\theta_k \in \Theta_k$ by an extended association map $\gamma_k : \mathbb{L}_k \to \{-1{:}|Z_k|\}$ defined by

$$\gamma_k(\ell) = \begin{cases} \theta_k(\ell), & \text{if } \ell \in \mathcal{D}(\theta_k) \\ -1, & \text{otherwise} \end{cases}. \tag{47}$$

Let Γ_k denote the set of positive 1-1 maps from $\mathbb{L}_k$ to $\{-1{:}|Z_k|\}$, and (with a slight abuse of notation) denote the *live labels* of γ_k, i.e. the domain $\mathcal{D}(\theta_k)$, by

$$\mathcal{L}(\gamma_k) \triangleq \{\ell \in \mathbb{L}_k : \gamma_k(\ell) \geq 0\}.$$

Then for any $\gamma_k \in \Gamma_k$, we can recover $\theta_k \in \Theta_k$ by $\theta_k(\ell) = \gamma_k(\ell)$ for each $\ell \in \mathcal{L}(\gamma_k)$. Hence, there is a bijection between Θ_k and Γ_k, and $\theta_{1:k}$ can be completely represented by $\gamma_{1:k}$.

Starting with the initial prior $\boldsymbol{\pi}_0(\mathbf{X}_0) = \delta_0[\mathcal{L}(\mathbf{X}_0)]$, and iteratively applying Proposition 8, the posterior at time k is:

$$\boldsymbol{\pi}_{0:k}(\mathbf{X}_{0:k}) \propto \Delta(\mathbf{X}_{0:k}) \sum_{\gamma_{0:k}} w_{0:k}^{(\gamma_{0:k})}\delta_{\mathcal{L}(\gamma_{0:k})}[\mathcal{L}(\mathbf{X}_{0:k})]\left[\tau_{0:k}^{(\gamma_{0:k}\circ\mathcal{L}(\cdot))}(\cdot)\right]^{\mathbf{X}_{0:k}} \tag{48}$$

where $\mathcal{L}(\gamma_0) \triangleq \emptyset$,

$$w_{0:k}^{(\gamma_{0:k})} = w_{0:k-1}^{(\gamma_{0:k-1})} 1_{\Gamma_k}(\gamma_k) 1_{\mathcal{F}(\mathbb{B}_k \uplus \mathcal{L}(\gamma_{k-1}))}(\mathcal{L}(\gamma_k))[\eta_{k|k-1}^{(\gamma_{0:k}(\cdot))}(\cdot)]^{\mathbb{B}_k \uplus \mathcal{L}(\gamma_{k-1})} \tag{49}$$

$$\eta_{k|k-1}^{(j_{s(\ell):k})}(\ell) \tag{50}$$
$$= \begin{cases} \bar{\Lambda}_{B,k}^{(j_k)}(\ell), & \ell \in \mathbb{B}_k, j_k \geq 0 \\ \bar{\Lambda}_{S,k|k-1}^{(j_{s(\ell):k})}(\ell), & \ell \in \mathbb{L}_{k-1}, j_k \geq 0 \\ Q_{B,k}(\ell), & \ell \in \mathbb{B}_k, j_k < 0 \\ \bar{Q}_{S,k-1}^{(j_{s(\ell):k-1})}(\ell), & \ell \in \mathbb{L}_{k-1}, j_k < 0 \end{cases}$$

$$\tau_{0:k}^{(j_{s(\ell):k})}(x_{s(\ell):t(\ell)}, \ell) \tag{51}$$
$$= \begin{cases} \frac{\Lambda_{B,k}^{(j_k)}(x_k,\ell)}{\bar{\Lambda}_{B,k}^{(j_k)}(\ell)}, & s(\ell) = k \\ \frac{\Lambda_{S,k|k-1}^{(j_k)}(x_k|x_{k-1},\ell)\tau_{0:k-1}^{(j_{s(\ell):k-1})}(x_{s(\ell):k-1},\ell)}{\bar{\Lambda}_{S,k|k-1}^{(j_{s(\ell):k})}(\ell)}, & t(\ell) = k > s(\ell) \\ \frac{Q_{S,k-1}(x_{k-1},\ell)\tau_{0:k-1}^{(j_{s(\ell):k-1})}(x_{s(\ell):k-1},\ell)}{\bar{Q}_{S,k-1}^{(j_{s(\ell):k-1})}(\ell)}, & t(\ell) = k-1 \\ \tau_{0:t(\ell)}^{(j_{s(\ell):t(\ell)})}(x_{s(\ell):t(\ell)},\ell), & t(\ell) < k-1, \end{cases}$$

$$\bar{\Lambda}_{S,i|i-1}^{(j_{s(\ell):i})}(\ell) \tag{52}$$
$$= \int \Lambda_{S,i|i-1}^{(j_i)}(x_i|x_{i-1},\ell)\tau_{0:i-1}^{(j_{s(\ell):i-1})}(x_{s(\ell):i-1},\ell)dx_{s(\ell):i}$$

$$\bar{Q}_{S,i-1}^{(j_{s(\ell):i-1})}(\ell) \tag{53}$$
$$= \int Q_{S,t(\ell)}(x_{i-1},\ell)\tau_{0:i-1}^{(j_{s(\ell):i-1})}(x_{s(\ell):i-1},\ell)dx_{s(\ell):i-1}$$

Note that the weight (49) can be written explicitly as

$$w_{0:k}^{(\gamma_{0:k})} = \prod_{i=1}^{k} 1_{\Gamma_i}(\gamma_i) 1_{\mathcal{F}(\mathbb{B}_i \uplus \mathcal{L}(\gamma_{i-1}))}(\mathcal{L}(\gamma_i))[\eta_{i|i-1}^{(\gamma_{0:i}(\cdot))}(\cdot)]^{\mathbb{B}_i \uplus \mathcal{L}(\gamma_{i-1})}.$$

Also, instead of using $\tau_{0:i-1}^{(j_{s(\ell):i-1})}(\cdot,\ell)$ in (52) and (53), we only need its marginal

$$\tau_{i-1}^{(j_{s(\ell):i-1})}(x_{i-1},\ell) = \int \tau_{0:i-1}^{(j_{s(\ell):i-1})}(x_{s(\ell):i-1},\ell)dx_{s(\ell):i-2}.$$

Computing $\tau_{0:k}^{(j_{s(\ell):k})}(\cdot,\ell)$ and $\eta_{k|k-1}^{(j_{s(\ell):k})}(\ell)$ is discussed in Appendix VII-F.

The multi-scan GLMB (48) is completely parameterized by the components $(w_{0:k}^{(\gamma_{0:k})}, \tau_{0:k}^{(\gamma_{0:k})})$, and we seek (without exhaustive enumeration) components with significant weights. This is a multi-dimensional assignment problem, which is NP-hard for more than two scans [52]. Our solution is based on sampling $\gamma_{0:k}$'s from some discrete probability distribution π such that components with high weights are more likely to be chosen than those with low weights. A natural choice is to set

$$\pi^{(j)}(\gamma_j|\gamma_{0:j-1}) \propto 1_{\Gamma_j}(\gamma_j) 1_{\mathcal{F}(\mathbb{B}_j \uplus \mathcal{L}(\gamma_{j-1}))}(\mathcal{L}(\gamma_j))[\eta_{j|j-1}^{(\gamma_{0:j}(\cdot))}(\cdot)]^{\mathbb{B}_j \uplus \mathcal{L}(\gamma_{j-1})} \tag{54}$$

with $\mathcal{L}(\gamma_j) = \emptyset$, so that

$$\pi(\gamma_{1:k}) = \prod_{j=1}^{k} \pi^{(j)}(\gamma_j|\gamma_{0:j-1}) \propto w_{0:k}^{(\gamma_{0:k})} \tag{55}$$

Metropolis-Hasting MCMC (MH-MCMC) is a popular technique for sampling from complex distributions, and has been applied to solve the data association problem for multi-object tracking in [53]. We seek to minimize the L_1 error between an approximate multi-scan GLMB and the multi-object posterior, by sampling GLMB components from (55). MH-MCMC could take some time for a new sample to be accepted, depending on the proposal, not to mention the time it takes for the chain to converge. Designing a proposal to have high acceptance probability is still an open area of research. Furthermore, the actual distribution of the samples and the convergence time depend on the starting value. Usually, an MCMC simulation is divided into two parts: the pre-convergence samples, known as burn-ins, are discarded; and the post-convergence samples are used for inference [54]. The key issue is that there are no bounds on the burn-in time nor reliable techniques for determining when convergence has occurred; see e.g. [54] and references therein.

The Gibbs sampler is a computationally efficient MCMC algorithm, in which proposed samples are always accepted [55], [56]. Further, for approximating multi-scan GLMB, the distribution of the samples is not relevant. Regardless of their distribution, all distinct samples will reduce the L_1 approximation error. However, Gibbs sampling requires the conditionals of (55), to be easily computed and sampled.

In the following, we present two techniques for sampling from (55). The first, detailed in subsection IV-B, is based on sampling from the factors (54), i.e., $\gamma_j \sim \pi^{(j)}(\cdot|\gamma_{0:j-1})$, for $j = 1:k$. The second, detailed in subsection IV-C, is a full Gibbs sampler with (55) as the stationary distribution.

B. Sampling from the Factors

Sampling from (54) using the Gibbs sampler [55], [56] involves constructing a Markov chain where a new state γ_j' is generated from state γ_j by sampling the values of $\gamma_j'(\ell_n)$, $\ell_n \in \{\ell_{1:|\mathbb{L}_j|}\} \triangleq \mathbb{L}_j$ from the distribution $\pi_n^{(j)}$ given by

$$\pi_n^{(j)}(\alpha|\gamma_j'(\ell_{1:n-1}), \gamma_j(\ell_{n+1:|\mathbb{L}_j|}), \gamma_{0:j-1}) \propto \pi^{(j)}(\gamma_j'(\ell_{1:n-1}), \alpha, \gamma_j(\ell_{n+1:|\mathbb{L}_j|})|\gamma_{0:j-1})$$

where

$$\gamma_j(\ell_{u:v}) \triangleq [\gamma_j(\ell_u), ..., \gamma_j(\ell_v)],$$
$$\gamma_j(\ell_{\bar{n}}) \triangleq [\gamma_j(\ell_{1:n-1}), \gamma_j(\ell_{n+1:|\mathbb{L}_j|})].$$

For a valid γ_j, i.e. $\pi^{(j)}(\gamma_j|\gamma_{0:j-1}) > 0$, it is necessary that $1_{\mathcal{F}(\mathbb{B}_j \uplus \mathcal{L}(\gamma_{j-1}))}(\mathcal{L}(\gamma_j)) = 1$, i.e. $\mathcal{L}(\gamma_j) \subseteq \mathbb{B}_j \uplus \mathcal{L}(\gamma_{j-1})$. This amounts to disregarding any γ_j that takes on a non-negative value outside $\mathbb{B}_j \uplus \mathcal{L}(\gamma_{j-1})$, and consider only those that take on -1 everywhere outside of $\mathbb{B}_j \uplus \mathcal{L}(\gamma_{j-1})$. In this case

$$\pi_n^{(j)}(\gamma_j(\ell_n)|\gamma_j(\ell_{\bar{n}}), \gamma_{0:j-1}) \propto 1_{\Gamma_j}(\gamma_j)\eta_{j|j-1}^{(\gamma_{0:j}(\ell_n))}(\ell_n)$$

for $\ell_n \in \{\ell_{1:|\mathbb{B}_j \uplus \mathcal{L}(\gamma_{j-1})|}\} \triangleq \mathbb{B}_j \uplus \mathcal{L}(\gamma_{j-1})$. Further, applying Proposition 3 of [19] gives:

$$\pi_n^{(j)}(\gamma_j(\ell_n)|\gamma_j(\ell_{\bar{n}}),\gamma_{0:j-1}) \propto \eta_{j|j-1}^{(\gamma_{0:j}(\ell_n))}(\ell_n) M^{(\gamma_j(\ell_{\bar{n}}))}(\gamma_j(\ell_n))$$

where

$$M^{(S)}(\alpha) = \begin{cases} 1, & \alpha \leq 0 \\ (1 - 1_S(\alpha)), & \alpha > 0 \end{cases}.$$

Hence, to generate γ_j' from a valid γ_j, we set $\gamma_j'(\ell) = -1$ for all $\ell \in \mathbb{L}_j - \mathbb{B}_j \uplus \mathcal{L}(\gamma_{j-1})$ and sample $\gamma_j'(\ell_n)$ for $\ell_n \in \{\ell_{1:|\mathbb{B}_j \uplus \mathcal{L}(\gamma_{j-1})|}\}$ from

$$\pi_n^{(j)}(\alpha|\gamma_j'(\ell_{1:n-1}), \gamma_j(\ell_{n+1:|\mathbb{L}_j|}), \gamma_{0:j-1}) \propto \eta_{j|j-1}^{(\gamma_{0:j-1}(\ell_n),\alpha)}(\ell_n) M^{(\gamma_j'(\ell_{1:n-1}), \gamma_j(\ell_{n+1:|\mathbb{L}_j|}))}(\alpha). \tag{56}$$

Note that in implementation, we only need the values of γ_j on $\mathbb{B}_j \uplus \mathcal{L}(\gamma_{j-1})$. The pseudo code for sampling from (55) by sampling from the factors is given in Algorithm 1.

Algorithm 1a: SampleJointFactors

- input: R (no. samples)
- output: $G_{0:k}$

initialize $G_0 := (\gamma_0, w_0, \tau_0)$;
for $j = 1:k$
  $[G_{0:j}^{(r)}]_{r=1}^{R} := \mathsf{SampleFactors}(G_{0:j-1}, R)$; $G_{0:j} := G_{0:j}^{(R)}$;
end

Algorithm 1b: SampleFactors

- input: $G_{0:k-1} = (\gamma_{0:k-1}, w_{0:k-1}, \tau_{0:k-1})$; R (no. samples)
- output: $[G_{0:k}^{(r)}]_{r=1}^{R} = [(\gamma_{0:k}^{(r)}, w_{0:k}^{(r)}, \tau_{0:k}^{(r)})]_{r=1}^{R}$

$P_k := |\mathbb{B}_k \uplus \mathcal{L}(\gamma_{k-1})|$; $M_k := |Z_k|$; $c := [-1{:}M_k]$;
$\gamma_k := \mathsf{zeros}(P_k, 1)$; (or any valid state)
for $r = 1:R$
  $\gamma_k' := [\,]$;
  for $n = 1:P_k$
    for $\alpha = -1:M_k$
      $\varkappa(\alpha) := \pi_n^{(k)}(\alpha|\gamma_k'(\ell_{1:n-1}), \gamma_k(\ell_{n+1:P_k}), \gamma_{0:k-1})$; via (56)
    end
    $\gamma_k'(\ell_n) \sim \mathsf{Categorical}(c, \varkappa)$; $\gamma_k' := [\gamma_k', \gamma_k'(\ell_n)]$;
  end
  $\gamma_k := \gamma_k'$; $\gamma_{0:k} := [\gamma_{0:k-1}, \gamma_k]$;
  compute $w_{0:k}, \tau_{0:k}$ from $\gamma_{0:k}$ via (49), (51);
  $G_{0:k}^{(r)} := (\gamma_{0:k}, w_{0:k}, \tau_{0:k})$;
end

This approach ensures that the sample $\gamma_{1:k}$ is a valid association history. However, to guarantee that $\gamma_{1:k}$ is a sample from (55), it is necessary to run each Gibbs sampler for sufficiently long at each time $j \in \{1:k\}$, to ensure that γ_j is a sample from $\pi^{(j)}(\cdot|\gamma_{0:j-1})$. Nonetheless, sampling from the factors can be used to generate a good starting point for the full Gibbs sampler.

C. Gibbs Sampling

Sampling from (55) using the Gibbs sampler [55], [56] involves constructing a Markov chain where a new state $\gamma_{1:k}'$ is generated from state $\gamma_{1:k}$ by sampling the values of $\gamma_j'(\ell_n)$, $j = 1:k$, $\ell_n \in \{\ell_1, ..., \ell_{|\mathbb{L}_j|}\}$ according to the conditional distribution $\pi_{j,n}$ defined by

$$\pi_{j,n}(\alpha|\overbrace{\gamma_{0:j-1}'}^{\text{past}}, \overbrace{\gamma_j'(\ell_{1:n-1})}^{\text{current (processed)}}, \overbrace{\gamma_j(\ell_{n+1:|\mathbb{L}_j|})}^{\text{current (unprocessed)}}, \overbrace{\gamma_{j+1:k}}^{\text{future}}) \propto \pi(\gamma_{0:j-1}', \gamma_j'(\ell_{1:n-1}), \alpha, \gamma_j(\ell_{n+1:|\mathbb{L}_j|}), \gamma_{j+1:k}).$$

We also use the notation $\gamma_{\bar{j}} \triangleq (\gamma_{0:j-1}, \gamma_{j+1:k})$ to denote the past and future association maps.

Observe from (54) and (55) that for a valid $\gamma_{1:k}$, i.e. $\pi(\gamma_{1:k}) > 0$, it is necessary that $1_{\Gamma_i}(\gamma_i) = 1$ (i.e. γ_i is positive 1-1), and $1_{\mathcal{F}(\mathbb{B}_i \uplus \mathcal{L}(\gamma_{i-1}))}(\mathcal{L}(\gamma_i)) = 1$ (i.e. dead labels at $i-1$ cannot be live at i, or equivalently, a live label at i cannot be dead at $i-1$) for $i = 1:k$. Thus, in addition to being positive 1-1, consecutive elements of a valid $\gamma_{1:k}$ must be such that dead labels remains dead at the next time. Closed form expressions for the conditionals are given in the following Proposition (see Appendix VII-G for proof).

Proposition 10. *Suppose $\gamma_j : \mathbb{L}_j \to \{-1{:}|Z_j|\}$, $j \in \{1{:}k\}$, is an element of a valid association history $\gamma_{1:k}$, and let*

$$\eta_{j,n}^{(\gamma_{\bar{j}})}(\alpha) \triangleq \prod_{i=j}^{t(\ell_n)} \eta_{i|i-1}^{(\gamma_{0:j-1}(\ell_n), \alpha, \gamma_{j+1:i}(\ell_n))}(\ell_n)$$

$$M_{\beta}^{(S)}(\alpha) \triangleq \begin{cases} \delta_{\beta}[\alpha], & \alpha < 0 \\ 1, & \alpha = 0 \\ (1 - 1_S(\alpha)), & \alpha > 0 \end{cases}$$

Then, for $\ell_n \in \{\ell_{1:|\mathbb{B}_j \uplus \mathcal{L}(\gamma_{j-1})|}\} \triangleq \mathbb{B}_j \uplus \mathcal{L}(\gamma_{j-1})$,

$$\pi_{j,n}(\gamma_j(\ell_n)|\gamma_j(\ell_{\bar{n}}), \gamma_{\bar{j}}) \propto \eta_{j,n}^{(\gamma_{\bar{j}})}(\gamma_j(\ell_n)) M_{\gamma_{\min\{j+1,k\}}(\ell_n)}^{(\gamma_j(\ell_{\bar{n}}))}(\gamma_j(\ell_n)) \quad (57)$$

and for $\ell_n \in \{\ell_{|\mathbb{B}_j \uplus \mathcal{L}(\gamma_{j-1})|+1:|\mathbb{L}_j|}\} \triangleq \mathbb{L}_j - \mathbb{B}_j \uplus \mathcal{L}(\gamma_{j-1})$

$$\pi_{j,n}(\gamma_j(\ell_n)|\gamma_j(\ell_{\bar{n}}), \gamma_{\bar{j}}) = \delta_{-1}[\gamma_j(\ell_n)] \delta_{\gamma_{\min\{j+1,k\}}(\ell_n)}[\gamma_j(\ell_n)]. \quad (58)$$

To generate γ_j', from a valid $\gamma_{1:k}$, we sample $\gamma_j'(\ell_n)$, $\ell_n \in \{\ell_{1:|\mathbb{B}_j \uplus \mathcal{L}(\gamma_{j-1})|}\}$ from

$$\pi_{j,n}(\alpha|\gamma_{0:j-}', \gamma_j'(\ell_{1:n-1}), \gamma_j(\ell_{n+1:|\mathbb{L}_j|}), \gamma_{j+1:k}) \propto \eta_{j,n}^{(\gamma_{0:j-1}', \gamma_{j+1:k})}(\alpha) M_{\gamma_{\min\{j+1,k\}}(\ell_n)}^{(\gamma_j'(\ell_{1:n-1}), \gamma_j(\ell_{n+1:|\mathbb{L}_j|}))}(\alpha) \quad (59)$$

and set $\gamma_j'(\ell_n) = -1$ for the remaining ℓ_n. This last step is omitted in actual implementation and it is understood that γ_j' is negative outside of $\{\ell_{1:|\mathbb{B}_j \uplus \mathcal{L}(\gamma_{j-1})|}\}$. The Gibbs sampler with stationary distribution (55) is given in Algorithm 2.

Algorithm 2: MultiScanGibbs

- input: $G_{0:k} = (\gamma_{0:k}, w_{0:k}, \tau_{0:k})$; T (no. samples)
- output: $[G_{0:k}^{(t)}]_{t=1}^{T}$

for $t = 1:T$
  for $j = 1:k$
    $P_j := |\mathbb{B}_j \uplus \mathcal{L}(\gamma_{j-1})|$; $M_j := |Z_j|$; $c := [-1{:}M_j]$; $\gamma_j' := [\,]$;
    for $n = 1:P_j$
      for $\alpha = -1:M_j$
        $\varkappa(\alpha) := \pi_{j,n}(\alpha|\gamma_{0:j-1}', \gamma_j'(\ell_{1:n-1}), \gamma_j(\ell_{n+1:P_j}), \gamma_{j+1:k})$;
        via (57)
      end
      $\gamma_j'(\ell_n) \sim \mathsf{Categorical}(c, \varkappa)$; $\gamma_j' := [\gamma_j'; \gamma_j'(\ell_n)]$;
    end
  $\gamma_j := \gamma_j'$; $\gamma_{0:j} := [\gamma_{0:j-1}, \gamma_j]$;
  compute $w_{0:k}, \tau_{0:k}$ from $\gamma_{0:j}$ via (49), (51);
  end
  $G_{0:k}^{(t)} := (\gamma_{0:k}, w_{0:k}, \tau_{0:k})$;
end

Starting with a valid association history, it follows from Proposition 10 that all iterates of the Gibbs sampler (Algorithm 2) are also valid association histories. The value $\eta_{j,n}^{(\gamma_{0:j-1}', \gamma_{j+1:k})}(\alpha)$ can be interpreted as the unnormalized

probability that the trajectory with label ℓ_n has the sequence of measurements with indices $\gamma'_{0:j-1}(\ell_n), \alpha, \gamma_{j+1:k}(\ell_n)$. The value $M^{(\gamma'_j(\ell_{1:n-1}),\gamma_j(\ell_{n+1:|\mathbb{L}_j|}))}_{\gamma_{\min\{j+1,k\}}(\ell_n)}(\alpha)$ is simply a binary mask which ensures that if ℓ_n is alive at the next time, $j+1$, then it cannot be dead at the current time j, or that if ℓ_n is alive at the current time j, then it cannot take on a measurement assigned to other labels at the current time. The product of these values is thus the unnormalized conditional probability that the trajectory with label ℓ_n has measurement index α at the current time, given $\gamma'_{0:j-1}, \gamma'_j(\ell_{1:n-1}), \gamma_j(\ell_{n+1:|\mathbb{L}_j|}), \gamma_{j+1:k}$, i.e. all of the other labels and associations from all other times. Notice (58) implies that if label ℓ_n is dead at the previous time $j-1$, then it must remain dead for all subsequent times.

Let $P_j = |\mathbb{B}_j \uplus \mathcal{L}(\gamma_{j-1})|$ and $M_j = |Z_j|$ then Algorithm 2 has complexity $\mathcal{O}(T \cdot \sum_{j=1}^{k} P_j^2 M_j)$, where T is the number of iterations of the Markov chain. Thus if $\bar{P} = \max_{j\in\{1:k\}} P_j$ and $\bar{M} = \max_{j\in\{1:k\}} M_j$ then indicatively Algorithm 2 has complexity $\mathcal{O}(kT\bar{P}^2\bar{M})$.

Remark: Algorithm 2 can also be implemented as a block Gibbs sampler. Instead of sampling every element of γ'_j from the conditionals (59), we draw the entire γ'_j from

$$\pi_j(\gamma'_j|\gamma'_{0:j-1}, \gamma_{j+1:k}) \propto \pi(\gamma'_{0:j-1}, \gamma'_j, \gamma_{j+1:k}).$$

Proposition 11. *Starting from any valid initial state, the Gibbs sampler defined by the conditionals (57) converges to the target distribution (55) at an exponential rate. More concisely, let π^j denote the jth power of the transition kernel, then*

$$\max_{\gamma_{1:k},\gamma'_{1:k}\in\Gamma_k} (|\pi^j(\gamma'_{1:k}|\gamma_{1:k}) - \pi(\gamma'_{1:k})|) \leq (1-2\beta)^{\lfloor \frac{j}{h} \rfloor},$$

where, $h = k+1$, $\beta \triangleq \min_{\gamma_{1:k},\gamma'_{1:k}\in\Gamma_k} \pi^h(\gamma'_{1:k}|\gamma_{1:k}) > 0$ is the least likely h-step transition probability.

The proof follows along the same lines as Proposition 4 of [19], with the 2-step transition probability replaced by the $(k+1)$-step transition probability. Instead of going from one arbitrary state of the chain to another via the all-zeros state in 2 steps as in [19], we go to the all-negative state (consisting of all -1) in k steps or less, and from this state to the other state in one additional step.

For batch smoothing, Algorithm 1b can be used to initialize the chain and Algorithm 2 can be used to generate samples from (55). The pseudocode for batch smoothing is given in Algorithm 3, where we enumerate the sum over $\gamma_{0:k}$ in (48) as a sum over $\{\gamma^{(h)}_{0:k}\}_{h=1}^{H}$ and use the shorthand $w^{(h)}_{0:k} = w^{(\gamma^{(h)}_{0:k})}_{0:k}$, $\tau^{(h)}_{0:k}(\cdot) = \tau^{(\gamma^{(h)}_{0:k}\circ\mathcal{L}(\cdot))}_{0:k}(\cdot)$. Recall that one of the proposed estimators is based on the most significant $\gamma_{1:k}$. In this case, the multi-scan Gibbs sampler can be used in a simulated annealing setting to find the best $\gamma_{1:k}$. The complexity of Algorithm 3 is consequently $\mathcal{O}(kT\bar{P}^2\bar{M})$.

To perform smoothing-while-filtering which propagates (48) recursively, Algorithm 1a can be used to propose a new ensemble of γ_k on-the-fly, and Algorithm 2 can be used to generate an ensemble of significant $\gamma_{0:k}$. Algorithm 4 presents the steps of a possible implementation of a smoothing-while-filtering iteration. Due the parallelizability of the for loops, the time complexity of Algorithm 4 is also $\mathcal{O}(kT\bar{P}^2\bar{M})$.

Algorithm 3: Batch

- input: $R; T$
- output: $[G^{(h)}_{0:k}]_{h=1}^{H}$

$G_{0:k} := \mathsf{SampleJointFactors}(R)$;
$[G^{(t)}_{0:k}]_{t=1}^{\tilde{T}} := \mathsf{Unique}(\mathsf{MultiScanGibbs}(G_{0:k}, T))$;
keep H best $[G^{(h)}_{0:k}]_{h=1}^{H}$;

Algorithm 4: Smoothing-while-Filtering

- input: $[G^{(h)}_{0:k-1}]_{h=1}^{H_{k-1}}; [T^{(h)}]_{h=1}^{H_{k-1}}; T$
- output: $[G^{(h)}_{0:k}]_{h=1}^{H_k}$

for $h = 1 : H_{k-1}$
  $[G^{(h,t)}_{0:k}]_{t=1}^{\tilde{T}^{(h)}} := \mathsf{Unique}(\mathsf{SampleFactors}(G^{(h)}_{0:k-1}, T^{(h)}))$;
end
keep $\bar{H}_k$ best $[G^{(h)}_{0:k}]_{h=1}^{\bar{H}_k}$;
for $h = 1 : \bar{H}_k$
  $[G^{(h,t)}_{0:k}]_{t=1}^{T} := \mathsf{MultiScanGibbs}(G^{(h)}_{0:k}, T)$;
end
$[G^{(h)}_{0:k}]_{h=1}^{\tilde{H}_k} := \mathsf{Unique}([G^{(h,t)}_{0:k}]_{h,t=(1,1)}^{(\bar{H}_k,T)})$;
keep H_k best $[G^{(h)}_{0:k}]_{h=1}^{H_k}$;
normalize weights $[w^{(h)}_{0:k}]_{h=1}^{H_k}$;

V. Numerical Experiments

A. Smoothing vs Filtering

This subsection demonstrates the performance of the multi-scan GLMB smoother proposed in Section IV. For benchmarking against GLMB filtering, we adopt the scenario in [19] Section IV(A)[2], a summary of which is given here. The duration is 100 time steps over which the number of objects is unknown and varies with time due to births and deaths. Births occur around times $1, 10, 60, 70$ (with respectively 3, 4, 2, 2 births) and multiple deaths occur around times $30, 50$ (2 deaths at both times). The 3 objects born at the beginning of the scenario cross at the origin around time 20, and another two pairs of objects cross, respectively on the left and right of the horizontal axis around time 40. A peak number of 9 objects occur simultaneously towards the end of the scenario.

Individual object dynamics and observations are linear Gaussian. The kinematic state of each object is a 4D state vector $[p_x, v_x, p_y, v_y]$ of 2D position and velocity, which follows a constant velocity model with a sampling period of $1s$ and process noise standard deviation $\sigma_\nu = 5m/s^2$. The survival probability P_S for each object is 0.99. Objects are born according to an LMB model with parameters $\{(r_{B,k}(\ell_i), p_{B,k}(\ell_i))\}_{i=1}^{3}$, where $\ell_i = (k,i)$, $r_{B,k}(\ell_i) = 0.04$, and $p_B(x,\ell_i) = \mathcal{N}(x; m_B^{(i)}, P_B)$ with

$$m_B^{(1)} = [0,0,100,0]^T, \quad m_B^{(2)} = [-100,0,-100,0]^T,$$
$$m_B^{(3)} = [100,0,-100,0]^T, \quad P_B = \mathrm{diag}([10,10,10,10]^T)^2.$$

[2]Errata: the conference version [44] used the scenario from [57] and not the scenario from [19] as originally quoted.

Observations are 2D position vectors $[z_x, z_y]$ on the region $[-1000, 1000]m \times [-1000, 1000]m$ with Gaussian noise standard deviation $\sigma_\varepsilon = 10m$. The detection probability $P_D = 0.66$ is considerably lower than the original value of 0.88 in [19]. Clutter follows a Poisson model with a uniform intensity $\kappa_k(z) = 1.93 \times 10^{-5}\ m^{-2}$ on the observation region, resulting in an average count of 77 false alarms per scan, which is higher than the original value of 66 in [19].

The multi-scan GLMB smoother (Algorithm 4) is run with a maximum $H = 1000$ components and for $T = 100$ iterations of the Markov chain, which is initialized by sampling from the factors (Algorithm 1a). To improve efficiency, we employed a block Gibbs sampling strategy. The single-scan GLMB filter [19] is also run with a maximum of 1000 components.

Figure 2 plots the GLMB smoother estimate and the true tracks in the x-y plane. Observe that the smoother initiates and terminates all pertinent tracks. By design of the smoother, there are no fragmented tracks. The estimates of the individual states and labels are consistent. In the presence of low detection probability and high clutter rate, there is however some increase in the positional errors during object crossings. The corresponding output for the GLMB filter is shown in Figure 3. In comparison there is a significant incidence of false, dropped and broken tracks, and even where a track has been declared, the positional errors are noticeably larger.

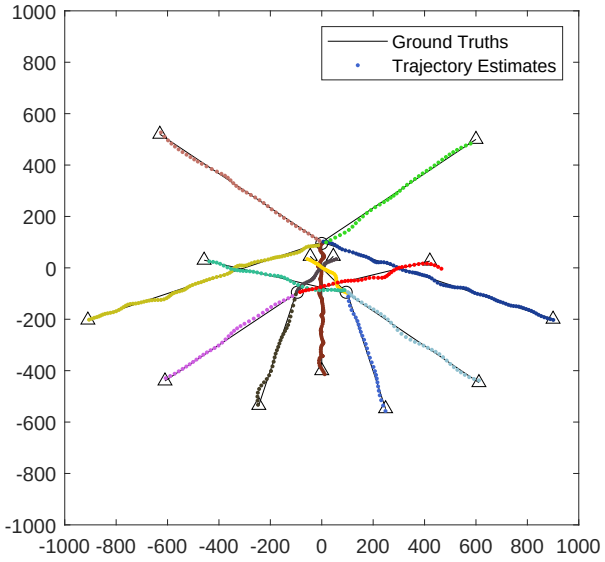

Fig. 2. Ground truths and estimated tracks from multi-scan GLMB smoothing. Starting and stopping positions are indicated with ◯ and △ respectively.

At each time, the smoother updates the entire multi-object history. Thus, for each multi-object trajectory estimate, it is important to consider the tracking errors at every instant of this history. Figure 4 plots the OSPA [58] and OSPA$^{(2)}$ [59] errors of the smoother's (and filter's) final estimate against time. 100 Monte Carlo trials are used with OSPA parameters $c = 100m$, and $p = 1$. The OSPA$^{(2)}$ error at time k is computed over a 10-scan window ending at time k and assesses tracking performance over this 10-scan window (see also [59], [60]). Due to the smoother's ability to correct earlier estimates, its

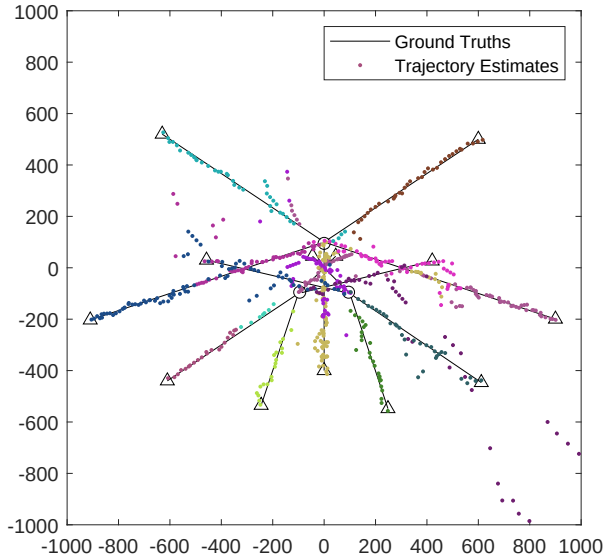

Fig. 3. Ground truths and estimated tracks for single-scan GLMB filtering. Starting and stopping positions are indicated with ◯ and △ respectively.

error is significantly below the filtering error for the entire duration of the estimate, albeit at higher computational cost.

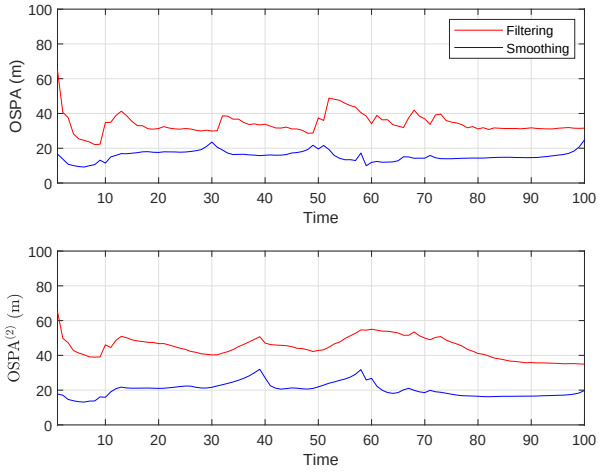

Fig. 4. OSPA and OSPA$^{(2)}$ error curves for the final estimates from the GLMB filter and multi-scan smoother over 100 Monte Carlo runs.

Remark: The latest technique for large scale multi-dimensional problems can handle up to 5 dimensions with 20 measurements per dimension, which translates to a linear program with 3.2 million binary variables [52]. Our 100-scan smoothing window example involves 100 dimensions with about 80 measurements per dimension, demonstrates the scalability of the proposed multi-dimensional assignment solution. In most multi-object tracking applications, such large smoothing windows are not needed. The longer the smoothing window, the more accurate the estimates, but the longer the computation time. A good trade-off for a multi-object a tracker is to use a moving smoothing window with shorter length.

B. Posterior Statistics Demonstration

This subsection considers a severe scenario to stress-test the multi-scan GLMB model and illustrate its capability to provide useful information even when tracking results are not reliable. The setting is a cell-biology application, where the user is interested in cell lifetime, birth rate, death rate, and cell migration patterns. The scenario duration is 1000 mins with data arriving at 10-min intervals. To simulate variations in cell lifetimes: at the 1st min 4 cells appear and live for 100 mins; at the 200th min another 4 cells appear and live for 200 mins; at the 500th min another 4 cells appear and live for 400 mins. The cells are initiated at the centres of the 4 quadrants of the region, with a common speed towards the origin and terminate before reaching the origin.

The cells have survival probability of 0.95, and their states are 4D vectors (planar position/velocity) that follow a constant velocity model with sampling period $\Delta = 10$*mins*, and process noise with standard deviation $\sigma_\nu = 0.01mm/\Delta^2$. The birth model is an LMB with parameters $\{(r_{B,k}(\ell_i), p_{B,k}(\ell_i))\}_{i=1}^{4}$, where $r_{B,k}(\ell_i) = 0.03$ and $p_B(x, \ell_i) = \mathcal{N}(x; m_B^{(i)}, P_B)$ with

$$m_B^{(1)} = [5, 0, 5, 0]^T, \quad m_B^{(2)} = [5, 0, -5, 0]^T,$$
$$m_B^{(3)} = [-5, 0, -5, 0]^T, \quad m_B^{(3)} = [-5, 0, 5, 0]^T,$$

and $P_B = \mathrm{diag}([0.15, 0.15, 0.15, 0.15]^T)^2$. Observations are 2D positions on the region $[-10, 10]mm \times [-10, 10]mm$, with additive Gaussian noise standard deviation $\sigma_\varepsilon = 0.3mm$. Clutter is uniform Poisson with a rate of 0.3 per scan, and the detection probability is 0.33. In practice data with such low detection probability will be discarded, we only use this scenario to demonstrate the capability of the proposed model. The multi-scan GLMB smoother is run with the same H, T and block sampling strategy as in the previous subsection.

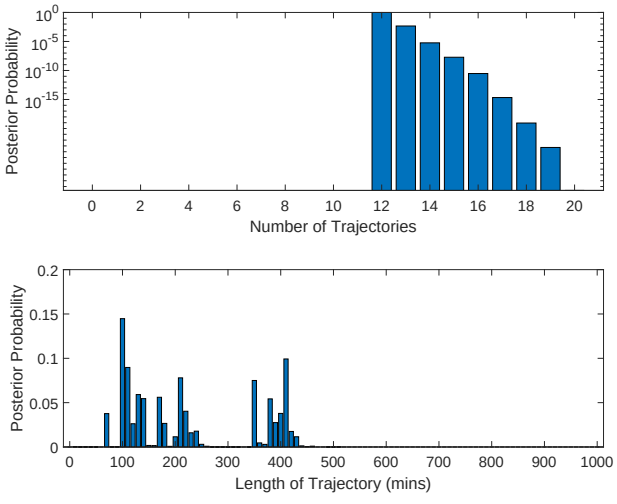

Fig. 5. Posterior distributions of number and lengths of trajectories.

Figure 5 shows the posterior distributions of the number of trajectories and cell lifetimes. Observe that the smoother correctly estimates 12 trajectories (with negligible modes at 13, ..., and 19), and the 3 modes of cell lifetimes (100, 200, and 400 mins). Note that shorter lifetime estimates have larger uncertainty. From Figure 6, showing the cardinality distributions of births and deaths in time, note that the smoother correctly identifies the instances of 4 births (at 1, 200, and 500 mins), and 4 deaths (at 100, 400, and 900 mins). There is however considerable uncertainty in the estimates of the death times due to the high uncertainty in the data.

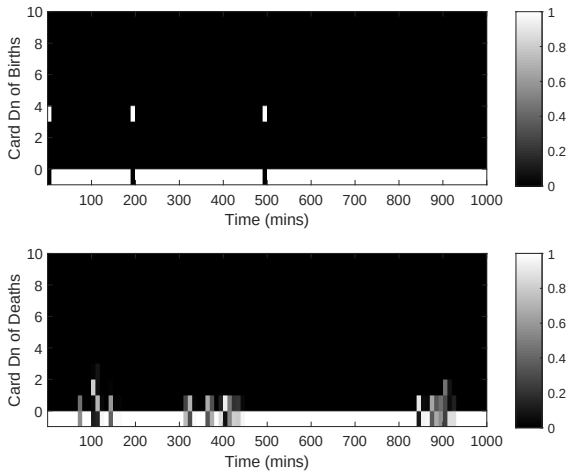

Fig. 6. Posterior cardinality distributions of births and deaths at each time.

Figure 7 shows the smoothed unlabeled PHD (or first moment) in the velocity space at 100, 200, 500, and 900 mins. The first 3 plots correctly confirm the presence of 4 modes of velocity (radially through the centre). The last plot correctly confirms the onset of cell deaths. Even with some uncertainty in the drift, these plots indicate overall migration of the cells diagonally across the region.

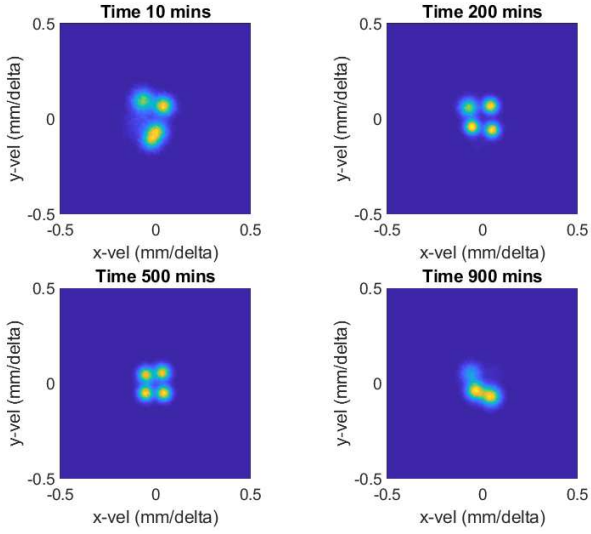

Fig. 7. Smoothed PHD in velocity space at various times.

VI. Conclusions

By introducing a multi-scan version of the GLMB model, we developed a multi-scan version of the GLMB filter to per-

form multi-object smoothing. We showed that computing the multi-scan GLMB posterior with minimal L_1-error (from its exact value) requires solving a multi-dimensional assignment problem with very high dimensions. Further, we developed an efficient and highly parallelizable algorithm for solving such multi-dimensional assignment problems using Gibbs sampling, and subsequently a novel multi-object smoothing-while-filtering algorithm. Numerical multi-object tracking examples demonstrated that the proposed algorithm significantly improves tracking performance as well as eliminating track fragmentation, a problem often found in multi-object filters. In addition, statistical characterization of variables/parameters pertaining to the underlying objects, can provide useful information, even for severe scenarios where multi-object trajectory estimates are no longer meaningful.

VII. Appendix

A. Properties of Multi-Scan Exponentials

To present relevant properties of multi-scan exponentials, we introduce some useful partitionings for the labels of the multi-object state sequence $\mathbf{X}_{j:k}$. Given a time i in $\{j:k\}$, a label $\ell \in \cup_{r=j}^{k}\mathcal{L}(\mathbf{X}_r)$ is alive at i iff $\ell \in \mathcal{L}(\mathbf{X}_i)$, terminates at $t(\ell) < i$ (before i) iff $\ell \in \mathcal{L}(\mathbf{X}_t) - \mathcal{L}(\mathbf{X}_{t+1})$, and born at time $s(\ell) > i$ (after i) iff $\ell \in \mathcal{L}(\mathbf{X}_s) \cap \mathbb{B}_s$. The set of labels in $\mathbf{X}_{j:k}$ can be partitioned into labels terminated before i, live labels at i, and labels born after i, i.e.

$$\bigcup_{r=j}^{k}\mathcal{L}(\mathbf{X}_r) = \overleftarrow{\mathcal{L}(\mathbf{X}_i)} \uplus \mathcal{L}(\mathbf{X}_i) \uplus \overrightarrow{\mathcal{L}(\mathbf{X}_i)} \tag{60}$$

where

$$\begin{aligned}\overleftarrow{\mathcal{L}(\mathbf{X}_i)} &\triangleq \{\ell \in \cup_{r=j}^{k}\mathcal{L}(\mathbf{X}_r) : t(\ell) < i\} \\ &= \bigcup_{r=j}^{i}\mathcal{L}(\mathbf{X}_r) - \mathcal{L}(\mathbf{X}_i) = \biguplus_{t=j}^{i-1}(\mathcal{L}(\mathbf{X}_t) - \mathcal{L}(\mathbf{X}_{t+1})) \quad (61)\\ \overrightarrow{\mathcal{L}(\mathbf{X}_i)} &= \{\ell \in \cup_{r=j}^{k}\mathcal{L}(\mathbf{X}_r) : s(\ell) > i\} \\ &= \bigcup_{r=i}^{k}\mathcal{L}(\mathbf{X}_r) - \mathcal{L}(\mathbf{X}_i) = \biguplus_{s=i+1}^{k}\mathcal{L}(\mathbf{X}_s) \cap \mathbb{B}_s. \quad (62)\end{aligned}$$

When $i = k$, $\overrightarrow{\mathcal{L}(\mathbf{X}_k)} = \emptyset$ and the set of labels in $\mathbf{X}_{j:k}$ can be partitioned into labels terminated before k and live labels at k, i.e. (61) becomes

$$\textstyle\bigcup_{r=j}^{k}\mathcal{L}(\mathbf{X}_r) = \overleftarrow{\mathcal{L}(\mathbf{X}_k)} \uplus \mathcal{L}(\mathbf{X}_k). \tag{63}$$

In addition, if $j = k-1$, then $\overleftarrow{\mathcal{L}(\mathbf{X}_k)} = \mathcal{L}(\mathbf{X}_{k-1}) - \mathcal{L}(\mathbf{X}_k)$, and using the decomposition $\mathcal{L}(\mathbf{X}_k) = (\mathcal{L}(\mathbf{X}_{k-1}) \cap \mathcal{L}(\mathbf{X}_k)) \uplus (\mathbb{B}_k \cap \mathcal{L}(\mathbf{X}_k))$, the set of labels in $\mathbf{X}_{k-1:k}$ can be partitioned into labels terminated at $k-1$, labels survived to k, and labels born at k, i.e. (63) becomes

$$\begin{aligned}&\textstyle\bigcup_{r=k-1}^{k}\mathcal{L}(\mathbf{X}_r) = \quad (64)\\ &(\mathcal{L}(\mathbf{X}_{k-1}) - \mathcal{L}(\mathbf{X}_k)) \uplus (\mathcal{L}(\mathbf{X}_{k-1}) \cap \mathcal{L}(\mathbf{X}_k)) \uplus (\mathbb{B}_k \cap \mathcal{L}(\mathbf{X}_k)).\end{aligned}$$

When $i = j$, $\overleftarrow{\mathcal{L}(\mathbf{X}_j)} = \emptyset$ and the set of labels in $\mathbf{X}_{j:k}$ can be partitioned into live labels at j and labels born after j, i.e. (62) becomes

$$\textstyle\bigcup_{r=j}^{k}\mathcal{L}(\mathbf{X}_r) = \mathcal{L}(\mathbf{X}_j) \uplus \overrightarrow{\mathcal{L}(\mathbf{X}_j)}. \tag{65}$$

B. Proof of Lemma 1

Parts (i) and (ii) follows straight from the definition of multi-scan exponential.

To prove (iii), noting from (60) that the set of labels $\cup_{r=j}^{k}\mathcal{L}(\mathbf{X}_r)$ can be partitioned into those terminated before i, alive at i, and born after i, we partition the set $\mathbf{X}_{j:k}$ of trajectories accordingly, i.e.

$$\begin{aligned}\mathbf{X}_{j:k} &= \{\mathbf{x}_{s(\ell):t(\ell)}^{(\ell)}: \ell \in \cup_{r=j}^{k}\mathcal{L}(\mathbf{X}_r)\} \\ &= \{\mathbf{x}_{s(\ell):t(\ell)}^{(\ell)}: \ell \in \overleftarrow{\mathcal{L}(\mathbf{X}_i)}\} \uplus \{\mathbf{x}_{s(\ell):t(\ell)}^{(\ell)}: \ell \in \mathcal{L}(\mathbf{X}_i)\} \\ &\quad \uplus \{\mathbf{x}_{s(\ell):t(\ell)}^{((\ell)}: \ell \in \overrightarrow{\mathcal{L}(\mathbf{X}_i)}\}.\end{aligned}$$

Hence, using (ii) gives

$$\begin{aligned}[h]^{\mathbf{X}_{j:k}} &= [h]^{\{\mathbf{x}_{s(\ell):t(\ell)}^{(\ell)}:\ell\in\overleftarrow{\mathcal{L}(\mathbf{X}_i)}\}}\,[h]^{\{\mathbf{x}_{s(\ell):t(\ell)}^{(\ell)}:\ell\in\mathcal{L}(\mathbf{X}_i)\}} \\ &\quad\times\ [h]^{\{\mathbf{x}_{s(\ell):t(\ell)}^{(\ell)}:\ell\in\overrightarrow{\mathcal{L}(\mathbf{X}_i)}\}}. \quad (66)\end{aligned}$$

and setting i to k, and i to j we have

$$[h]^{\mathbf{X}_{j:k}} = [h]^{\{\mathbf{x}_{s(\ell):t(\ell)}^{(\ell)}:\ell\in\overleftarrow{\mathcal{L}(\mathbf{X}_k)}\}}\,[h]^{\{\mathbf{x}_{s(\ell):k}^{(\ell)}:\ell\in\mathcal{L}(\mathbf{X}_k)\}}, \tag{67}$$

$$[h]^{\mathbf{X}_{j:k}} = [h]^{\{\mathbf{x}_{j:t(\ell)}^{(\ell)}:\ell\in\mathcal{L}(\mathbf{X}_j)\}}\,[h]^{\{\mathbf{x}_{s(\ell):t(\ell)}^{(\ell)}:\ell\in\overrightarrow{\mathcal{L}(\mathbf{X}_j)}\}}. \tag{68}$$

Using (67) and (68), we partition $[g]^{\mathbf{X}_{j:i}}$ and $[h]^{\mathbf{X}_{i:k}}$, and then combine them as follows

$$\begin{aligned}[g]^{\mathbf{X}_{j:i}}[h]^{\mathbf{X}_{i:k}} &= [g]^{\{\mathbf{x}_{s(\ell):t(\ell)}^{(\ell)}:\ell\in\overleftarrow{\mathcal{L}(\mathbf{X}_i)}\}}\,[g]^{\{\mathbf{x}_{s(\ell):i}^{(\ell)}:\ell\in\mathcal{L}(\mathbf{X}_i)\}} \\ &\quad\times [h]^{\{\mathbf{x}_{i:t(\ell)}^{(\ell)}:\ell\in\mathcal{L}(\mathbf{X}_i)\}}\,[h]^{\{\mathbf{x}_{s(\ell):t(\ell)}^{(\ell)}:\ell\in\overrightarrow{\mathcal{L}(\mathbf{X}_i)}\}} \\ &= [g]^{\{\mathbf{x}_{s(\ell):t(\ell)}^{(\ell)}:\ell\in\overleftarrow{\mathcal{L}(\mathbf{X}_i)}\}}\prod_{\ell\in\mathcal{L}(\mathbf{X}_i)} g(\mathbf{x}_{s(\ell):i}^{(\ell)}) \\ &\quad\times \prod_{\ell\in\mathcal{L}(\mathbf{X}_i)} h(\mathbf{x}_{i:t(\ell)}^{(\ell)})\,[h]^{\{\mathbf{x}_{s(\ell):t(\ell)}^{(\ell)}:\ell\in\overrightarrow{\mathcal{L}(\mathbf{X}_i)}\}} \\ &= [g\odot h]^{\{\mathbf{x}_{s(\ell):t(\ell)}^{(\ell)}:\ell\in\overleftarrow{\mathcal{L}(\mathbf{X}_i)}\}}[g\odot h]^{\{\mathbf{x}_{s(\ell):t(\ell)}^{(\ell)}:\ell\in\mathcal{L}(\mathbf{X}_i)\}} \\ &\quad\times [g\odot h]^{\{\mathbf{x}_{s(\ell):t(\ell)}^{(\ell)}:\ell\in\overrightarrow{\mathcal{L}(\mathbf{X}_i)}\}} \\ &= [g\odot h]^{\mathbf{X}_{j:k}}.\end{aligned}$$

C. Proof of Proposition 2 (Multi-Object Transition)

Using (64) to partition $\mathcal{L}(\mathbf{X}_{k-1}) \cup \mathcal{L}(\mathbf{X}_k)$ into disappearing labels at time $k-1$, surviving labels at time k, and new born labels at time k, and noting that

$$\begin{aligned}\{\mathbf{x}_k^{(\ell)}: \ell \in \mathbb{B}_k \cap \mathcal{L}(\mathbf{X}_k)\} &= \{\mathbf{x}_k^{(\ell)}: s(\ell) = k\} \\ \{\mathbf{x}_{k-1:k}^{(\ell)} : \ell \in \mathcal{L}(\mathbf{X}_{k-1}) \cap \mathcal{L}(\mathbf{X}_k)\} &= \{\mathbf{x}_{k-1:k}^{(\ell)} : t(\ell) = k\} \\ \{\mathbf{x}_{k-1}^{(\ell)} : \ell \in \mathcal{L}(\mathbf{X}_{k-1}) - \mathcal{L}(\mathbf{X}_k)\} &= \{\mathbf{x}_{k-1}^{(\ell)} : t(\ell) = k-1\}\end{aligned}$$

we have

$$\begin{aligned}\mathbf{X}_{k-1:k} &\equiv \{\mathbf{x}_{s(\ell):t(\ell)}^{(\ell)} : \ell \in \mathcal{L}(\mathbf{X}_{k-1}) \cup \mathcal{L}(\mathbf{X}_k)\} \\ &= \{\mathbf{x}_{k-1}^{(\ell)} : t(\ell) = k-1\} \uplus \{\mathbf{x}_{k-1:k}^{(\ell)} : t(\ell) = k\} \\ &\quad \uplus \{\mathbf{x}_k^{(\ell)} : s(\ell) = k\}.\end{aligned}$$

Let $\mathbf{x}_k^{(\ell)} = (x_k^{(\ell)}, \ell)$ to denote the element of the multi-object state $\mathbf{X}_k$ at time k, with label $\ell \in \mathcal{L}(\mathbf{X}_k)$, then the multi-object transition density given in [16], [18] can be rewritten as

$$
\begin{aligned}
&\mathbf{f}_{k|k-1}(\mathbf{X}_k|\mathbf{X}_{k-1}) \\
&= \Delta(\mathbf{X}_k)1_{\mathcal{F}(\mathbb{B}_k \uplus \mathcal{L}(\mathbf{X}_{k-1}))}(\mathcal{L}(\mathbf{X}_k))Q_{B,k}^{\mathbb{B}_k-\mathcal{L}(\mathbf{X}_k)} \\
&\times \prod_{\ell \in \mathcal{L}(\mathbf{X}_{k-1})-\mathcal{L}(\mathbf{X}_k)} Q_{S,k-1}(x_{k-1}^{(\ell)},\ell) \prod_{\ell \in \mathbb{B}_k \cap \mathcal{L}(\mathbf{X}_k)} P_{B,k}(\ell) f_{B,k}(x_k^{(\ell)},\ell) \\
&\times \prod_{\ell \in \mathcal{L}(\mathbf{X}_{k-1}) \cap \mathcal{L}(\mathbf{X}_k)} P_{S,k-1}(x_{k-1}^{(\ell)},\ell) f_{S,k|k-1}(x_k^{(\ell)}|x_{k-1}^{(\ell)},\ell) \\
&= \Delta(\mathbf{X}_k)1_{\mathcal{F}(\mathbb{B}_k \uplus \mathcal{L}(\mathbf{X}_{k-1}))}(\mathcal{L}(\mathbf{X}_k))Q_{B,k}^{\mathbb{B}_k-\mathcal{L}(\mathbf{X}_k)} \\
&\times \left[\phi_{k-1:k}\right]^{\{\mathbf{x}_{k-1}^{(\ell)}:t(\ell)=k-1\}} \left[\phi_{k-1:k}\right]^{\{\mathbf{x}_{k-1:k}^{(\ell)}:t(\ell)=k\}} \\
&\times \left[\phi_{k-1:k}\right]^{\{\mathbf{x}_k^{(\ell)}:s(\ell)=k\}} \\
&= \Delta(\mathbf{X}_k)1_{\mathcal{F}(\mathbb{B}_k \uplus \mathcal{L}(\mathbf{X}_{k-1}))}(\mathcal{L}(\mathbf{X}_k))Q_{B,k}^{\mathbb{B}_k-\mathcal{L}(\mathbf{X}_k)} \left[\phi_{k-1:k}\right]^{\mathbf{X}_{k-1:k}}
\end{aligned}
$$

where the last step follows from (68).

D. Proof of Proposition 3

Using the δ-form we have

$$
\begin{aligned}
&\int f(\mathcal{L}(\mathbf{X}_{j:k}))\boldsymbol{\pi}(\mathbf{X}_{j:k})\delta\mathbf{X}_{j:k} \\
&= \int f(\mathcal{L}(\mathbf{X}_{j:k})) \sum_{\xi, I_{j:k}} w^{(\xi)}(I_{j:k})\delta_{j:k}[\mathcal{L}(\mathbf{X}_{j:k})][p^{(\xi)}]^{\mathbf{X}_{j:k}}\delta\mathbf{X}_{j:k} \\
&= \sum_{\xi, I_{j:k}} f(I_{j:k})w^{(\xi)}(I_{j:k}) \int \delta_{j:k}[\mathcal{L}(\mathbf{X}_{j:k})][p^{(\xi)}]^{\mathbf{X}_{j:k}}\delta\mathbf{X}_{j:k} \\
&= \sum_{\xi, I_{j:k}} f(I_{j:k})w^{(\xi)}(I_{j:k}) \prod_{\ell \in I_{j:k}} \int p^{(\xi)}(x_{s(\ell):t(\ell)}^{(\ell)},\ell)dx_{s(\ell):t(\ell)}^{(\ell)} \quad (69) \\
&= \sum_{\xi, I_{j:k}} f(I_{j:k})w^{(\xi)}(I_{j:k}),
\end{aligned}
$$

where (69) follows from Lemma D below.

Lemma D: For a function h taking trajectories to the reals, with $h(\cdot,\ell)$ integrable for each $\ell \in \cup_{i=j}^{k} I_i \equiv I_{j:k}$

$$
\int \delta_{I_{j:k}}[\mathcal{L}(\mathbf{X}_{j:k})][h]^{\mathbf{X}_{j:k}}\delta\mathbf{X}_{j:k} = \prod_{\ell \in I_{j:k}} \int h(x_{s(\ell):t(\ell)}^{(\ell)},\ell)dx_{s(\ell):t(\ell)}^{(\ell)}.
$$

Proof: For $g : \mathcal{F}(\mathbb{X}\times\mathbb{L}) \to \mathbb{R}$ and $I = \{i_1, ..., i_{|I|}\} \subseteq \mathbb{L}$,

$$
\begin{aligned}
&\int \delta_I[\mathcal{L}(\mathbf{X})]g(\mathbf{X})\delta\mathbf{X} \\
&= \sum_{n=0}^{\infty} \sum_{l_{1:n}} \frac{1}{n!} \int \delta_I[\{l_1, ..., l_n\}]g(\{(l_1, x_1), ..., (l_n, x_n)\})dx_{1:n} \\
&= \int g(\{(i_1, x_1), ..., (i_{|I|}, x_{|I|})\})dx_{1:|I|} \quad (70)
\end{aligned}
$$

For $g : \mathcal{F}(\mathbb{X}\times\mathbb{L}_j)\times\cdots\times\mathcal{F}(\mathbb{X}\times\mathbb{L}_k) \to \mathbb{R}$, and $I_t = \{i_{t,1}, ..., i_{t,|I_t|}\} \subseteq \mathbb{L}_t, t = j, ..., k$,

$$
\begin{aligned}
&\int \delta_{I_{j:k}}[\mathcal{L}(\mathbf{X}_{j:k})]g(\mathbf{X}_{j:k})\delta\mathbf{X}_j \\
&\quad = \delta_{I_{j+1:k}}[\mathcal{L}(\mathbf{X}_{j+1:k})] \int \delta_{I_j}[\mathcal{L}(\mathbf{X}_j)]g(\mathbf{X}_j, \mathbf{X}_{j+1:k})\delta\mathbf{X}_j \\
&\quad = \delta_{I_{j+1:k}}[\mathcal{L}(\mathbf{X}_{j+1:k})] \\
&\quad \times \int g(\{(i_{j,1}, x_{j,1}), ..., (i_{j,|I_j|}, x_{j,|I_j|})\}, \mathbf{X}_{j+1:k})dx_{j,1:|I_j|}
\end{aligned}
$$

where the last line follows from (70). Further, iterating for $j+1$, ..., k

$$
\begin{aligned}
&\int \delta_{I_{j:k}}[\mathcal{L}(\mathbf{X}_{j:k})]g(\mathbf{X}_{j:k})\delta\mathbf{X}_{j:k} \\
&\quad = \int ... \int g(\{(i_{j,1}, x_{j,1}), ..., (i_{j,N_j}, x_{j,|I_j|})\}, ..., \\
&\qquad \{(i_{k,1}, x_{k,1}), ..., (i_{k,N_k}, x_{k,|I_k|})\})dx_{j,1:|I_j|} \cdots dx_{k,1:|I_k|}
\end{aligned}
$$

Setting $g(\mathbf{X}_{j:k}) = [h]^{\mathbf{X}_{j:k}}$ yields

$$
\begin{aligned}
&\int \delta_{I_{j:k}}[\mathcal{L}(\mathbf{X}_{j:k})][h]^{\mathbf{X}_{j:k}}\delta\mathbf{X}_{j:k} \\
&= \int \delta_{I_{j:k}}[\mathcal{L}(\mathbf{X}_{j:k})] \prod_{\ell \in I_{j:k}} h(\mathbf{x}_{s(\ell):t(\ell)}^{(\ell)})\delta\mathbf{X}_{j:k} \\
&= \int ... \int \prod_{\ell \in I_{j:k}} h(\mathbf{x}_{s(\ell):t(\ell)}^{(\ell)})dx_{j,1:|I_j|} \cdots dx_{k,1:|I_k|} \\
&= \int ... \int \prod_{\ell \in I_{j:k}} h(x_{s(\ell):t(\ell)}^{(\ell)},\ell)dx_{j,1} \cdots dx_{j,|I_j|} \cdots dx_{k,1} \cdots dx_{k,|I_k|} \\
&= \prod_{\ell \in I_{j:k}} \int h(x_{s(\ell):t(\ell)}^{(\ell)},\ell)dx_{s(\ell):t(\ell)}^{(\ell)}
\end{aligned}
$$

where the last step follows from regrouping $dx_{j,1} \cdots dx_{j,|I_j|} \cdots dx_{k,1} \cdots dx_{k,|I_k|}$ to $\prod_{\ell \in I_{j:k}} dx_{s(\ell):t(\ell)}^{(\ell)}$.

E. Proof of Proposition 8 (Multi-Scan GLMB Recursion)

Since $\boldsymbol{\pi}_{0:k-1}(\mathbf{X}_{0:k-1})$ is the multi-scan GLMB (36) with $j = 0$, substituting it into the posterior recursion (6), gives the multi-scan GLMB (38)-(41) with $j = 0$, i.e.

$$
\begin{aligned}
&\boldsymbol{\pi}_{0:k}(\mathbf{X}_{0:k}) \\
&\quad \propto g_k(Z_k|\mathbf{X}_k)\mathbf{f}_{k|k-1}(\mathbf{X}_k|\mathbf{X}_{k-1})\boldsymbol{\pi}_{0:k-1}(\mathbf{X}_{0:k-1}) \\
&\quad = g_k(Z_k|\mathbf{X}_{0:k}) \left(\mathbf{f}_{k|k-1}(\mathbf{X}_k|\mathbf{X}_{k-1})\boldsymbol{\pi}_{0:k-1}(\mathbf{X}_{0:k-1})\right) \\
&\quad \propto \Delta(\mathbf{X}_{0:k}) \sum_{\xi \in \Xi} \sum_{\theta_k \in \Theta_k} w_{0:k}^{(\xi,\theta_k)}(\mathcal{L}(\mathbf{X}_{0:k}))[p_{0:k}^{(\xi,\theta_k)}]^{\mathbf{X}_{0:k}}.
\end{aligned}
$$

To evaluate $p_{0:k}^{(\xi,\theta_k)}(\cdot,\ell)$ as per (39), note from (36) that

$$
\begin{aligned}
&p_{0:k}^{(\xi)}(x_{s(\ell):t(\ell)}^{(\ell)},\ell) \\
&= (p_{0:k-1}^{(\xi)} \odot \phi_{k-1:k})(x_{s(\ell):t(\ell)}^{(\ell)},\ell) \\
&= \begin{cases}
\phi_{k-1:k}(x_{s(\ell):t(\ell)}^{(\ell)},\ell), & s(\ell) > k-1 \\
p_{0:k-1}^{(\xi)}(x_{s(\ell):k-1}^{(\ell)},\ell)\phi_{k-1:k}(x_{k-1:t(\ell)}^{(\ell)},\ell), & s(\ell) \le k-1 \le t(\ell) \\
p_{0:k-1}^{(\xi)}(x_{s(\ell):t(\ell)}^{(\ell)},\ell), & t(\ell) < k-1
\end{cases} \\
&= \begin{cases}
P_{B,k}(\ell)f_{B,k}(x_k^{(\ell)},\ell), & s(\ell) = k \\
p_{0:k-1}^{(\xi)}(x_{s(\ell):k-1}^{(\ell)},\ell)P_{S,k-1}(x_{k-1}^{(\ell)},\ell) & \\
\times f_{S,k|k-1}(x_k^{(\ell)}|x_{k-1}^{(\ell)},\ell), & s(\ell) < t(\ell) = k \\
p_{0:k-1}^{(\xi)}(x_{s(\ell):k-1}^{(\ell)},\ell)Q_{S,k-1}(x_{k-1}^{(\ell)},\ell), & s(\ell) \le k-1 = t(\ell) \\
p_{0:k-1}^{(\xi)}(x_{s(\ell):t(\ell)}^{(\ell)},\ell), & t(\ell) < k-1
\end{cases}
\end{aligned}
$$

Moreover, multiplying by $\psi_{0:k,Z_k}^{(\theta_k(\ell))}(x_{s(\ell):t(\ell)}^{(\ell)},\ell)$,

$$
\begin{aligned}
&p_{0:k}^{(\xi)}(x_{s(\ell):t(\ell)}^{(\ell)},\ell)\psi_{0:k,Z_k}^{(\theta_k(\ell))}(x_{s(\ell):t(\ell)}^{(\ell)},\ell) \quad (71) \\
&= \begin{cases}
\Lambda_{B,k}^{(\theta_k(\ell))}(x_k^{(\ell)},\ell), & s(\ell) = k \\
p_{0:k-1}^{(\xi)}(x_{s(\ell):k-1}^{(\ell)},\ell)\Lambda_{S,k|k-1}^{(\theta_k(\ell))}(x_k^{(\ell)}|x_{k-1}^{(\ell)},\ell), & s(\ell) < t(\ell) = k \\
p_{0:k-1}^{(\xi)}(x_{s(\ell):k-1}^{(\ell)},\ell)Q_{S,k-1}(x_{k-1}^{(\ell)},\ell), & s(\ell) \le k-1 = t(\ell) \\
p_{0:k-1}^{(\xi)}(x_{s(\ell):t(\ell)}^{(\ell)},\ell), & t(\ell) < k-1
\end{cases}
\end{aligned}
$$

and integrating we have

$$
\bar{\psi}_{0:k,Z_k}^{(\xi,\theta_k)}(\ell) = \begin{cases}
\bar{\Lambda}_{B,k}^{(\theta_k(\ell))}(\ell), & s(\ell) = k \\
\bar{\Lambda}_{S,k|k-1}^{(\xi,\theta_k(\ell))}(\ell), & s(\ell) < t(\ell) = k \\
\bar{Q}_{S,k-1}^{(\xi)}(\ell), & s(\ell) \le k-1 = t(\ell) \\
1, & t(\ell) < k-1
\end{cases} \quad (72)
$$

Note that in the last step we used

$$\begin{aligned}
&\int \Lambda_{S,k}^{(\theta_k(\ell))}(x_k,\ell|x_{k-1}) p_{0:k-1}^{(\xi)}(x_{s(\ell):k-1},\ell) dx_{s(\ell):k} \\
&= \int \Lambda_{S,k}^{(\theta_k(\ell))}(x_k,\ell|x_{k-1}) p_{k-1}^{(\xi)}(x_{k-1},\ell) dx_{k-1:k} = \bar{\Lambda}_{S,k,}^{(\xi,\theta_k)}(\ell) \\
&\int Q_{S,k-1}(x_{k-1},\ell) p_{0:k-1}^{(\xi)}(x_{s(\ell):k-1},\ell) dx_{s(\ell):k-1} \\
&= \int Q_{S,k-1}(x_{k-1},\ell) p_{k-1}^{(\xi)}(x_{k-1},\ell) dx_{k-1} = \bar{Q}_{S,k-1}^{(\xi)}(\ell)
\end{aligned}$$

Hence, dividing (71) by (72) according to (39) gives (46).

Using the following equivalences: $s(\ell)=k$ iff $\ell \in \mathcal{D}(\theta_k) \cap \mathbb{B}_k$; $s(\ell) < t(\ell) = k$ iff $\ell \in \mathcal{D}(\theta_k) - \mathbb{B}_k$; $s(\ell) \leq k-1 = t(\ell)$ iff $\ell \in I_{k-1} - \mathcal{D}(\theta_k)$, we have

$$\begin{aligned}
&\left[\bar{\psi}_{0:k,Z_k}^{(\xi,\theta_k)}\right]^{I_{0:k}} Q_{B,k}^{\mathbb{B}_k - \mathcal{D}(\theta_k)} \\
&\quad = \left[\bar{\Lambda}_{B,k}^{(\theta_k(\cdot))}(\cdot)\right]^{\mathcal{D}(\theta_k)\cap\mathbb{B}_k} \left[\bar{\Lambda}_{S,k|k-1}^{(\xi,\theta_k(\cdot))}(\cdot)\right]^{\mathcal{D}(\theta_k)-\mathbb{B}_k} \\
&\quad \times \left[\bar{Q}_{S,k-1}^{(\xi)}(\cdot)\right]^{I_{k-1}-\mathcal{D}(\theta_k)} Q_{B,k}^{\mathbb{B}_k-\mathcal{D}(\theta_k)} \\
&\quad = \left[\omega_{k|k-1}^{(\xi,\theta_k)}\right]^{\mathbb{B}_k \uplus I_{k-1}} \qquad (73)
\end{aligned}$$

since $\mathcal{D}(\theta_k)\cap\mathbb{B}_k$, $\mathcal{D}(\theta_k)-\mathbb{B}_k$, $I_{k-1}-\mathcal{D}(\theta_k)$, and $\mathbb{B}_k - \mathcal{D}(\theta_k)$ form a partition of $\mathbb{B}_k \uplus I_{k-1}$.

Noting that for any $\theta_k \in \Theta_k$, $1_{\Theta_k(I_k)}(\theta_k) = \delta_{\mathcal{D}(\theta_k)}[I_k]$, and substituting (36), (73) into definition (41) we have

$$\begin{aligned}
w_{0:k}^{(\xi,\theta_k)}(I_{0:k}) &= 1_{\mathcal{F}(\mathbb{B}_k \uplus I_{k-1})}(\mathcal{D}(\theta_k)) \left[\omega_{k|k-1}^{(\xi,\theta_k)}\right]^{\mathbb{B}_k \uplus I_{k-1}} \\
&\quad \times w_{0:k-1}^{(\xi)}(I_{0:k-1}) \delta_{\mathcal{D}(\theta_k)}[I_k] \\
&= \omega_{0:k}^{(\xi,\theta_k)}(I_{0:k-1}) \delta_{\mathcal{D}(\theta_k)}[I_k].
\end{aligned}$$

F. Computing Multi-Scan GLMB Parameters

Under a linear Gaussian multi-object model:

$$\begin{aligned}
\psi_{k,\{z_{1:m}\}}^{(j)}(x,\ell) &= \begin{cases} \frac{P_{D,k}^{(\ell)} \mathcal{N}(z_j; H_k x, R_k)}{\kappa_k(z_j)}, & \text{if } j > 0 \\ Q_{D,k}^{(\ell)} & \text{if } j = 0 \end{cases} \\
P_{S,k-1}(\varsigma,\ell) &= P_{S,k-1}^{(\ell)}, \quad Q_{S,k-1}(\varsigma,\ell) = Q_{S,k-1}^{(\ell)} \\
f_{S.k|k-1}(x|\varsigma,\ell) &= \mathcal{N}(x; F_{k|k-1}\varsigma, Q_k) \\
P_{B,k}(\ell) &= P_{B,k}^{(\ell)}, \quad Q_{B,k}(\ell) = Q_{B,k}^{(\ell)} \\
f_{B,k}(x,\ell) &= \mathcal{N}(x; m_{B,k}^{(\ell)}, Q_{B,k}^{(\ell)})
\end{aligned}$$

where $\mathcal{N}(\cdot;\mu,\Sigma)$ denotes a Gaussian density with mean μ and covariance Σ, $F_{k|k-1}$ and H_k are the single-object transition and measurement matrices, Q_k and R_k are the process and measurement noise covariances, $m_{B,k}^{(\ell)}$ and $Q_{B,k}^{(\ell)}$ are the mean and covariance of any new state with label ℓ. It follows from (51) that the densities $\tau_{0:k}^{(j_{s(\ell):k})}(\cdot,\ell)$ are Gaussians. Further, the (canonical) multi-scan GLMB parameters, $\tau_{0:k}^{(j_{s(\ell):k})}(\cdot,\ell)$ and $\eta_{k|k-1}^{(j_{s(\ell):k})}(\ell)$ (required for the Gibbs sampler) can be computed recursively using the following standard results on joint and conditional Gaussians

$$\begin{aligned}
&\mathcal{N}(z; Hx, R)\mathcal{N}(x; m, P) \\
&= \mathcal{N}\left([x;z], \hat{\mu}(H,m), \hat{\Sigma}(H,R,P)\right) \\
&= \mathcal{N}\left(x; \mu(z,H,R,m,P), \Sigma(H,R,P)\right) q(z; H, R, m, P)
\end{aligned}$$

where

$$\begin{aligned}
\hat{\mu}(H,m) &\triangleq \begin{bmatrix} m \\ Hm \end{bmatrix}, \\
\hat{\Sigma}(H,R,P) &\triangleq \begin{bmatrix} P & PH^T \\ HP & R + HPH^T \end{bmatrix}, \\
\mu(z,H,R,m,P) &\triangleq m + PH^T(R+HPH^T)^{-1}(z-Hm), \\
\Sigma(H,R,P) &\triangleq P - PH^T(R+HPH^T)^{-1}HP, \\
q(z;H,R,m,P) &\triangleq \mathcal{N}\left(z; Hm, R + HPH^T\right).
\end{aligned}$$

Let d denotes the dimension of the single-object state space, and $\Pi_{j:k} \triangleq [0_{d,(k-j)d}, I_{d,d}]$, then these identities become:

$$\begin{aligned}
&\mathcal{N}(x_k; Fx_{k-1}, Q)\mathcal{N}(x_{j:k-1}; m_{j:k-1}, P_{j:k-1}) \\
&\quad = \mathcal{N}(x_k; F\Pi_{j:k-1}x_{j:k-1}, Q)\mathcal{N}(x_{j:k-1}; m_{j:k-1}, P_{j:k-1}) \\
&\quad = \mathcal{N}(x_{j:k}; \hat{m}_{j:k}, \hat{P}_{j:k}) \\
\hat{m}_{j:k} &= \hat{\mu}(F\Pi_{j:k-1}, m_{j:k-1}) \\
\hat{P}_{j:k} &= \hat{\Sigma}(F\Pi_{j:k-1}, Q, P_{j:k-1}), \\
&\mathcal{N}(z; Hx_k, R)\mathcal{N}(x_{j:k}; \hat{m}_{j:k}, \hat{P}_{j:k}) \\
&\quad = \mathcal{N}(z; H\Pi_{j:k}x_{j:k}, R)\mathcal{N}(x_{j:k}; \hat{m}_{j:k}, \hat{P}_{j:k}) \\
&\quad = \mathcal{N}\left(x_{j:k}; m_{j:k}, P_{j:k}\right) q(z; H\Pi_{j:k}, R, \hat{m}_{j:k}, \hat{P}_{j:k}) \\
m_{j:k} &= \mu(z, H\Pi_{j:k}, R, \hat{m}_{j:k}, \hat{P}_{j:k}) \\
P_{j:k} &= \Sigma(H\Pi_{j:k}, R, \hat{P}_{j:k}).
\end{aligned}$$

Hence, for the multi-scan GLMB parameters we have:

$$\begin{aligned}
&\tau_{0:k}^{(j_{s(\ell):k})}(x_{s(\ell):t(\ell)}, \ell) \\
&\qquad = \mathcal{N}(x_{s(\ell):t(\ell)}; m_{0:t(\ell)}^{(j_{s(\ell):t(\ell)})}(\ell), P_{0:t(\ell)}^{(j_{s(\ell):t(\ell)})}(\ell)) \\
&\eta_{k|k-1}^{(j_{s(\ell):k})}(\ell) = \begin{cases} P_{B,k}^{(\ell)} q_k^{(j_k)}(z_{j_k},\ell), & \ell \in \mathbb{B}_k, j_k \geq 0 \\ P_{S,k-1}^{(\ell)} q_k^{(j_{s(\ell):k})}(z_{j_k},\ell), & \ell \in \mathbb{L}_{k-1}, j_k \geq 0 \\ Q_{B,k}^{(\ell)}, & \ell \in \mathbb{B}_k, j_k < 0 \\ Q_{S,k-1}^{(\ell)}, & \ell \in \mathbb{L}_{k-1}, j_k < 0 \end{cases}
\end{aligned}$$

where

$$\begin{aligned}
&m_{0:k}^{(j_{s(\ell):k})}(\ell) = \\
&\qquad \begin{cases} \hat{m}_{0:k}^{(j_{s(\ell):k})}(\ell), & j_k = 0 \\ \mu(z_{j_k}, H_k\Pi_k^{(\ell)}, R_k, \hat{m}_{0:k}^{(j_{s(\ell):k})}(\ell), \hat{P}_{0:k}^{(j_{s(\ell):k})}(\ell)), & j_k > 0 \end{cases} \\
&P_{0:k}^{(j_{s(\ell):k})}(\ell) = \begin{cases} \hat{P}_{0:k}^{(j_{s(\ell):k})}(\ell), & j_k = 0 \\ \Sigma(H_k\Pi_k^{(\ell)}, R_k, \hat{P}_{0:k}^{(j_{s(\ell):k})}(\ell)), & j_k > 0 \end{cases} \\
&\hat{m}_{0:k}^{(j_{s(\ell):k})}(\ell) = \begin{cases} m_{B,k}^{(\ell)}, & s(\ell) = k \\ \hat{\mu}(F_{k|k-1}\Pi_{k-1}^{(\ell)}, m_{0:k-1}^{(j_{s(\ell):k-1})}(\ell)), & s(\ell) < k \end{cases} \\
&\hat{P}_{0:k}^{(j_{s(\ell):k})}(\ell) = \begin{cases} Q_{B,k}^{(\ell)}, & s(\ell) = k \\ \hat{\Sigma}(F_{k|k-1}\Pi_{k-1}^{(\ell)}, Q_k, P_{0:k-1}^{(j_{s(\ell):k-1})}(\ell)), & s(\ell) < k \end{cases} \\
&\Pi_j^{(\ell)} = \begin{bmatrix} 0_{d,(j-s(\ell))d}, & I_{d,d} \end{bmatrix} \\
&q_k^{(j_{s(\ell):k})}(z_{j_k},\ell) = \begin{cases} Q_{D,k}^{(\ell)}, & j_k = 0 \\ \frac{P_{D,k}^{(\ell)} q(z_{j_k}; H_k, R_k, \hat{m}_k^{(j_{s(\ell):k})}(\ell), \hat{P}_k^{(j_{s(\ell):k})}(\ell))}{\kappa_k(z_{j_k})}, & j_k > 0 \end{cases} \\
&\hat{m}_k^{(j_{s(\ell):k})}(\ell) = \Pi_k^{(\ell)} \hat{m}_{0:k}^{((j_{s(\ell):k})}(\ell), \\
&\hat{P}_k^{(j_{s(\ell):k})}(\ell) = \Pi_k^{(\ell)} \hat{P}_{0:k}^{((j_{s(\ell):k})}(\ell)(\Pi_k^{(\ell)})^T
\end{aligned}$$

For non-linear non-Gaussian models particle smoothing methods can be used to approximate $\tau_{0:k}^{(j_{s(\ell):k})}(\cdot,\ell)$ and subsequently $\eta_{k|k-1}^{(j_{s(\ell):k})}(\ell)$ by replacing integrals with sums.

G. Proof of Proposition 10

Using (55) and noting that we are only interested in the functional dependence of $\pi_{j,n}(\gamma_j(\ell_n)|\gamma_j(\ell_{\bar{n}}),\gamma_{\bar{j}})$ on $\gamma_j(\ell_n)$, we write

$$\begin{aligned}\pi_{j,n}(\gamma_j(\ell_n)|\gamma_j(\ell_{\bar{n}}),\gamma_{\bar{j}}) &\propto \pi(\gamma_{0:j-1},\gamma_j(\ell_n),\gamma_j(\ell_{\bar{n}}),\gamma_{j+1:k}) \\ &= \prod_{i=1}^{j-1}\pi^{(i)}(\gamma_i|\gamma_{0:i-1})\prod_{i=j}^{k}\pi^{(i)}(\gamma_i|\gamma_{0:i-1}) \\ &\propto \prod_{i=j}^{k}\pi^{(i)}(\gamma_i|\gamma_{0:i-1})\end{aligned}$$

since $\gamma_j(\ell_n)$ is not contained in any of the factors $\pi^{(i)}(\gamma_i|\gamma_{0:i-1})$, $i\in\{1{:}j-1\}$. Substituting (54) for the remaining factors gives

$$\begin{aligned}&\pi_{j,n}(\gamma_j(\ell_n)|\gamma_j(\ell_{\bar{n}}),\gamma_{\bar{j}}) \\ &\propto \prod_{i=j}^{k}1_{\Gamma_i}(\gamma_i)1_{\mathcal{F}(\mathbb{B}_i\uplus\mathcal{L}(\gamma_{i-1}))}(\mathcal{L}(\gamma_i))\prod_{i=j}^{k}\prod_{\ell\in\mathbb{B}_i\uplus\mathcal{L}(\gamma_{i-1})}\eta_{i|i-1}^{(\gamma_{0:i}(\ell))}(\ell) \\ &= \prod_{i=j}^{k}1_{\Gamma_i}(\gamma_i)1_{\mathcal{F}(\mathbb{B}_i\uplus\mathcal{L}(\gamma_{i-1}))}(\mathcal{L}(\gamma_i))\prod_{i=j}^{t(\ell_n)}\eta_{i|i-1}^{(\gamma_{0:i}(\ell_n))}(\ell_n) \\ &\times\left(\prod_{i=j}^{k}\prod_{\ell\in\mathbb{B}_i\uplus\mathcal{L}(\gamma_{i-1})-\{\ell_n\}}\eta_{i|i-1}^{(\gamma_{0:i}(\ell))}(\ell)\right) \\ &\propto 1_{\mathcal{F}(\mathbb{B}_j\uplus\mathcal{L}(\gamma_{j-1}))}(\mathcal{L}(\gamma_j))1_{\mathcal{F}(\mathbb{B}_{j+1}\uplus\mathcal{L}(\gamma_j))}(\mathcal{L}(\gamma_{j+1}))1_{\Gamma_j}(\gamma_j) \\ &\times\eta_{j,n}^{(\gamma_{\bar{j}})}(\gamma_j(\ell_n))\end{aligned}$$

where in the last step we aggregated all terms not involving $\gamma_j(\ell_n)$ into the normalizing constant.

The validity of $\gamma_{1:k}$ implies both $1_{\mathcal{F}(\mathbb{B}_j\uplus\mathcal{L}(\gamma_{j-1}))}(\mathcal{L}(\gamma_j))$ and $1_{\mathcal{F}(\mathbb{B}_{j+1}\uplus\mathcal{L}(\gamma_j))}(\mathcal{L}(\gamma_{j+1}))$ equal to 1, which means that the following conditions hold

$$\forall\ell\in\mathbb{L}_j-\mathbb{B}_j\uplus\mathcal{L}(\gamma_{j-1}),\gamma_j(\ell)=-1, \tag{74}$$

$$\forall\ell\in\mathbb{L}_j,\gamma_j(\ell)\geq 0 \text{ or } \gamma_{\min\{j+1,k\}}(\ell)=-1. \tag{75}$$

Violation of (74) means $1_{\mathcal{F}(\mathbb{B}_j\uplus\mathcal{L}(\gamma_{j-1}))}(\mathcal{L}(\gamma_j))=0$, because if there exist an $\ell\in\mathbb{L}_j-\mathbb{B}_j\uplus\mathcal{L}(\gamma_{j-1})$ such that $\gamma_j(\ell)\geq 0$, then $\mathcal{L}(\gamma_j)$ is not in $\mathbb{B}_j\uplus\mathcal{L}(\gamma_{j-1})$, i.e. $1_{\mathcal{F}(\mathbb{B}_j\uplus\mathcal{L}(\gamma_{j-1}))}(\mathcal{L}(\gamma_j))=0$.

Violation of (75) means $1_{\mathcal{F}(\mathbb{B}_{j+1}\uplus\mathcal{L}(\gamma_j))}(\mathcal{L}(\gamma_{j+1}))=0$, because (except for $j=k$) if there exist an $\ell\in\mathbb{L}_j$ such that $\gamma_j(\ell)<0$ and $\gamma_{j+1}(\ell)\geq 0$, then ℓ is not in $\mathcal{L}(\gamma_j)$ and hence $\mathcal{L}(\gamma_{j+1})$ (which contains ℓ) is not contained in $\mathbb{B}_{j+1}\uplus\mathcal{L}(\gamma_j)$, i.e. $1_{\mathcal{F}(\mathbb{B}_{j+1}\uplus\mathcal{L}(\gamma_j))}(\mathcal{L}(\gamma_{j+1}))=0$.

We consider $\pi_{j,n}(\gamma_j(\ell_n)|\gamma_j(\ell_{\bar{n}}),\gamma_{\bar{j}})$ for $\ell_n\in\{\ell_{1:|\mathbb{B}_j\uplus\mathcal{L}(\gamma_{j-1})|}\}$ first, and subsequently for $\ell_n\in\{\ell_{|\mathbb{B}_j\uplus\mathcal{L}(\gamma_{j-1})|+1:|\mathbb{L}_j|}\}$, if this set is non-empty.

For any $\ell_n\in\{\ell_{1:|\mathbb{B}_j\uplus\mathcal{L}(\gamma_{j-1})|}\}$, (74) holds, and we have either (i) $\gamma_j(\ell_n)\geq 0$ or $\gamma_{\min\{j+1,k\}}(\ell_n)=-1$; or (ii) its (logical) complement, i.e. $\gamma_j(\ell_n)<0$ and $\gamma_{\min\{j+1,k\}}(\ell_n)\geq 0$.

For case (i), since (75) also holds,

$$\begin{aligned}&\pi_{j,n}(\gamma_j(\ell_n)|\gamma_j(\ell_{\bar{n}}),\gamma_{\bar{j}}) \\ &\propto 1_{\Gamma_j}(\gamma_j)\eta_{j,n}^{(\gamma_{\bar{j}})}(\gamma_j(\ell_n)) \\ &\propto\begin{cases}\eta_{j,n}^{(\gamma_{\bar{j}})}(\gamma_j(\ell_n)), & \gamma_j(\ell_n)\leq 0\\ \eta_{j,n}^{(\gamma_{\bar{j}})}(\gamma_j(\ell_n))(1-1_{\gamma_j(\ell_{\bar{n}})}(\gamma_j(\ell_n))), & \gamma_j(\ell_n)>0\end{cases},\end{aligned}$$

where the last step invokes Proposition 3 of [19].

Case (ii) violates (75), hence $1_{\mathcal{F}(\mathbb{B}_{j+1}\uplus\mathcal{L}(\gamma_j))}(\mathcal{L}(\gamma_{j+1}))=0$ and consequently $\pi(\gamma_j(\ell_n)|\gamma_j(\ell_{\bar{n}}),\gamma_{\bar{j}})=0$.

Decomposing $\gamma_j(\ell_n)\leq 0$ into two cases $\gamma_j(\ell_n)=0$ and $\gamma_j(\ell_n)<0$, and combining the latter with case (ii) we have

$$\begin{aligned}&\pi_{j,n}(\gamma_j(\ell_n)|\gamma_j(\ell_{\bar{n}}),\gamma_{\bar{j}}) \\ &\propto\begin{cases}\eta_{j,n}^{(\gamma_{\bar{j}})}(\gamma_j(\ell_n))\delta_{\gamma_{\min\{j+1,k\}}(\ell_n)}[\gamma_j(\ell_n)], & \gamma_j(\ell_n)<0\\ \eta_{j,n}^{(\gamma_{\bar{j}})}(\gamma_j(\ell_n)), & \gamma_j(\ell_n)=0\\ \eta_{j,n}^{(\gamma_{\bar{j}})}(\gamma_j(\ell_n))(1-1_{\gamma_j(\ell_{\bar{n}})}(\gamma_j(\ell_n))), & \gamma_j(\ell_n)>0\end{cases},\end{aligned}$$

and hence (57).

For any $\ell_n\in\{\ell_{|\mathbb{B}_j\uplus\mathcal{L}(\gamma_{j-1})|+1:|\mathbb{L}_j|}\}$, any values other than $\gamma_j(\ell_n)=-1$ and $\gamma_{\min\{j+1,k\}}(\ell_n)=-1$ would violate (74) or (75). Either of these violations imply $1_{\mathcal{F}(\mathbb{B}_{j+1}\uplus\mathcal{L}(\gamma_j))}(\mathcal{L}(\gamma_{j+1}))=0$. Hence we have (58).

REFERENCES

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

[2] M. Briers, A. Doucet, and S. Maskell, "Smoothing algorithms for state-space models," *Ann. Inst. Stats. Math.*, 62(1):61–89, Springer Netherlands, 2010.

[3] A. Doucet, and A. M. Johansen, "A tutorial on particle filtering and smoothing: Fifteen years later," *Handbook of nonlinear filtering*, 12(3):656–704, 2009.

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

[5] A. Doucet, S. Godsill, and C. Andrieu, "On sequential Monte Carlo sampling methods for Bayesian filtering," *Stats. & Comp.*, 10(3):197–208, 2000.

[6] B. Ristic, S. Arulampalam, and N. Gordon, *Beyond the Kalman filter: Particle filters for tracking applications*. Artech House, 2004.

[7] R. Mahler, "Multitarget Bayes filtering via first-order multitarget moments," *IEEE Trans. Aerosp. Electron. Syst.*, 39(4):1152–1178, 2003.

[8] R. Mahler, "PHD filters of higher order in target number," *IEEE Trans. Aerosp. Electron. Syst.*, 43(4):1523–1543, 2007.

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

[10] R. Mahler, *Advances in Statistical Multisource-Multitarget Information Fusion*, Artech House, 2014.

[11] B.-T. Vo, B.-N. Vo, and A. Cantoni, "The cardinality balanced Multi-Target Multi-Bernoulli filter and its implementations," *IEEE Trans. Sig. Proc.*, 57(2):409–423, 2009.

[12] J. Correa, M., Adams, and C. Perez, "A Dirac delta mixture-based random finite set filter," *IEEE Intl. Conf. Cntrl, Aut. & Info. Sci.* (ICCAIS), pp. 231-238, Oct, 2015.

[13] J. L. Williams, "An efficient, variational approximation of the best fitting multi-Bernoulli filter," *IEEE Trans. Sig. Proc.*, 63(1): 258-273, 2015.

[14] A. García-Fernández, J. Williams, K. Granstrom, and L. Svensson "Poisson multi-Bernoulli mixture filter: direct derivation and implementation," *IEEE Trans. Aerosp. Electron. Syst.*, 2018.

[15] I. Schlangen, E.Delande, J. Houssineau, and D. Clark, "A second-order PHD filter with mean and variance in target number," *IEEE Trans. Sig. Proc.*, 66:48-63, 2018.

[16] B.-T. Vo and B.-N. Vo, "Labeled random finite sets and multi-object conjugate priors," *IEEE Trans. Sig. Proc.*, 61(13):3460–3475, 2013.

[17] T. Vu, B.-N. Vo and R.J. Evans, "A Particle Marginal Metropolis-Hastings Multi-target Tracker," *IEEE Trans. Sig. Proc.*, 62(15):3953–3964, 2014.

[18] B.-N. Vo, B.-T. Vo, and D. Phung, "Labeled random finite sets and the Bayes multi-target tracking filter," *IEEE Trans. Sig. Proc.*, 62(24):6554–6567, 2014.

[19] B.-N. Vo, B.-T. Vo, and H. Hoang, "An Efficient Implementation of the Generalized Labeled Multi-Bernoulli Filter," *IEEE Trans. Sig. Proc.,* 65(8):1975–1987, 2017.

[20] S. Reuter, B.-T. Vo, B.-N. Vo, and K. Dietmayer, "The labeled multi-Bernoulli filter," *IEEE Trans. Sig. Proc.*, 62(12):3246–3260, 2014.

[21] F. Papi, B.-N. Vo, B.-T. Vo, C. Fantacci, and M. Beard, "Generalized labeled multi-Bernoulli approximation of multi-object densities," *IEEE Trans. Sig. Proc.*, 63(20):5487–5497, 2015.

[22] C. Fantacci, and F. Papi, "Scalable multisensor multitarget tracking using the marginalized-GLMB density," *IEEE Sig. Proc. Lett.* 23(6):863-867, 2016.

[23] Z. Lu, W. Hu, and T. Kirubarajan, "Labeled random finite sets with moment approximation," *IEEE Trans. Sig. Proc.*, 65(13):3384-3398, 2017.

[24] S. Li, W. Yi, R. Hoseinnezhad, B. Wang, and L. Kong, "Multiobject Tracking for Generic Observation Model Using Labeled Random Finite Sets," *IEEE Trans. Sig. Proc.*, 66(2):368–383, 2018.

[25] F. Papi and D. Y. Kim, "A particle multi-target tracker for superpositional measurements using labeled random finite sets," *IEEE Trans. Sig. Proc.,* 63(16):4348-4358, 2015.

[26] M. Beard, B.-T. Vo, B.-N. Vo, and S. Arulampalam "Void probabilities and Cauchy-Schwarz divergence for Generalized Labeled multi-Bernoulli models," *IEEE Trans. Sig. Proc.*, 65(19):5047–5061, 2017.

[27] A.K. Gostar, R. Hoseinnezhad, T. Rathnayake, X. Wang, and A. Bab-Hadiashar, "Constrained Sensor Control for Labeled Multi-Bernoulli Filter Using Cauchy-Schwarz Divergence," *IEEE Sig. Proc. Lett.*, 24(9):1313–1317, 2017.

[28] S. Panicker, A.K. Gostar, A. Bab-Hadiashar, and R. Hoseinnezhad, "Accelerated Multi-Sensor Control for Selective Multi-Object Tracking," *IEEE Intl. Conf. Cntrl. Aut. & Info. Sci.* (ICCAIS), pp. 183-188, October, 2018.

[29] H. Deusch, S. Reuter, and K. Dietmayer, "The labeled multi-Bernoulli SLAM filter," *IEEE Sig. Proc. Lett.*, 22(10):1561-1565, 2015.

[30] D. Moratuwage, M. Adams, and F. Inostroza, "δ-Generalised Labelled Multi-Bernoulli Simultaneous Localisation and Mapping," *IEEE Intl. Conf. Cntrl. Aut. & Info. Sci.* (ICCAIS), pp. 175-182, October, 2018.

[31] C. Fantacci, B.-N. Vo, B.-T. Vo, G. Battistelli, and L. Chisci, "Robust fusion for multisensor multiobject tracking," *IEEE Sig. Proc. Lett.*, 25(5):640–644, 2018.

[32] X. Wang, A.K. Gostar, T. Rathnayake, B, Xu, and A. Bab-Hadiashar, "Centralized multiple-view sensor fusion using labeled multi-Bernoulli filters," *Signal Processing*, 150:75–84, 2018.

[33] A.K. Gostar,T. Rathnayake, A. Bab-Hadiashar, G. Battistelli, and L. Chisci, et. al. 'Centralized Multiple-View Information Fusion for Multi-Object Tracking Using Labeled Multi-Bernoulli Filters', *IEEE Intl. Conf. Cntrl, Aut. & Info. Sci.* (ICCAIS), October, 2018.

[34] S. Li, W. Yi, R. Hoseinnezhad, G. Battistelli, B, Wang, and L. Kong, "Robust distributed fusion with labeled random finite sets," *IEEE Trans. Sig. Proc.*, 66(2), 278-293, 2018.

[35] S. Li, G. Battistelli., L. Chisci, W. Yi, B. Wang, and L. Kong, "Computationally Efficient Multi-Agent Multi-Object Tracking with Labeled Random Finite Sets," *IEEE Trans. Sig. Proc.*, 67(1):260-275, 2019.

[36] P. Bunch, and S. Godsill, "Particle smoothing algorithms for variable rate models," *IEEE Trans. Sig. Proc.*, 61(7):663-1675, 2013.

[37] F. Papi, M. Bocquel, M. Podt, and Y. Boers, "Fixed-lag smoothing for Bayes optimal knowledge exploitation in target tracking," *IEEE Trans. Sig. Proc.*, 62(12):3143-3152, 2014.

[38] A. Finke, and S. Singh, "Approximate Smoothing and Parameter Estimation in High-Dimensional State-Space Models," *IEEE Trans. Sig. Proc.*, 65(22):5982-5994, 2017.

[39] J. Olsson, and J. Westerborn,."Efficient particle-based online smoothing in general hidden Markov models: the PaRIS algorithm," *Bernoulli*, 23(3):1951-1996, 2017.

[40] M. Gerber, and N. Chopin, "Convergence of sequential quasi-Monte Carlo smoothing algorithms," *Bernoulli*, 23(4B):2951-2987, 2017.

[41] S. G. Megason and S. E. Fraser, "Imaging in systems biology," *Cells*, 130:784-795, 2007.

[42] K. Duffy, C. Wellard, J. Markham, et al., "Activation-induced B cell fates are selected by intracellular stochastic competition", *Science*, 335(6066):338-341, 2012.

[43] M. Dowling, A. Kan, S. Heinzel, et al., "The stretched cell cycle model for proliferating lymphocytes", *Proc. Natl. Acad. Sci.,* 111(17):6377-6382, 2014.

[44] B.-T. Vo and B.-N. Vo, "Multi-Scan Generalized Labeled Multi-Bernoulli Filter," Proc. *21st. Conf. Inf. Fusion*, Cambridge, UK, 2018.

[45] B.-N. Vo, S. Singh, and A. Doucet, "Sequential Monte Carlo methods for multi-target filtering with random finite sets," *IEEE Trans. Aerosp. Electron. Syst.,*, 41(4):1224–1245, 2005.

[46] C. Andrieu, A. Doucet, and R. Holenstein, "Particle Markov chain Monte Carlo methods," *J. Royal Stat. Soc. B-Stat. Method.*, 72(3):269–342, 2010.

[47] P. Craciun, M. Ortner, and J. Zerubia, "Joint detection and tracking of moving objects using spatio-temporal marked point processes," *Proc. IEEE Winter Conf. App. Comp. Vision*, pp. 177–184 Jan. 2015.

[48] A.F. García-Fernández, L. Svensson, and M. Morelande, "Multiple target tracking based on sets of trajectories," *arXiv preprint* arXiv:1605.08163, 2018.

[49] M. Jacobsen, *Point Process Theory and Applications: Marked Point and Piecewise Deterministic Process*. Boston: Birkhauser, 2006.

[50] H. Hoang, B.-N. Vo, B.-T., Vo, and R. Mahler, "The Cauchy-Schwarz divergence for Poisson point processes," *IEEE Trans. Inf. Theory*, 61(8):4475- 4485, 2015.

[51] W. P. Pierskalla, Letter to the Editor—The Multidimensional Assignment Problem. *Operations Research*, 16(2):422-431, 1968.

[52] D. M. Nguyen, H.A.T. Le, and T. D. Pham, "Solving the Multidimensional Assignment Problem by a Cross-Entropy method," *J. Comb. Optim.,* 27:808–823, 2014.

[53] S. Oh, S. Russell, and S. Sastry, "Markov Chain Monte Carlo Data Association for Multi-Target Tracking," *IEEE Trans. Autom. Control,* 54(3):481–497, 2009.

[54] S. El Adlouni, A.-C. Favre, and B. Bobee, ""Comparison of methodologies to assess the convergence of Markov Chain Monte Carlo methods," *Comp. Stats. & Data Analysis*, 50(10):2685-2701, 2006.

[55] S. Geman and D. Geman, "Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images," *IEEE Trans. Pattern Anal. & Mach. Intell.*, 6(6):721–741, 1984.

[56] G. Casella and E. I. George, "Explaining the Gibbs sampler," *The American Statistician*, 46(3):167–174, 1992.

[57] B.-T. Vo, B.-N. Vo and A. Cantoni, "Analytic implementations of the Cardinalized Probability Hypothesis Density Filter," *IEEE Trans. Sig. Proc.*, 55(7 part 2):3553-3567, 2007.

[58] D. Schumacher, B.-T. Vo, and B.-N. Vo, "A consistent metric for performance evaluation of multi-object filters," *IEEE Trans. Sig. Proc.*, 56(8):3447–3457, 2008.

[59] M. Beard, B.-T. Vo, and B.-N. Vo, "A Solution for Large-Scale Multi-Object Tracking," *arXiv:1804.06622 [stat.CO],* Apr. 2018.

[60] M. Beard, B.-T. Vo, and B.-N. Vo, "OSPA$^{(2)}$: Using the OSPA metric to evaluate multi-target tracking performance," *IEEE Intl. Conf. Cntrl. Aut. & Info. Sci.* (ICCAIS), pp. 86-91, Nov. 2017.