

Multiple Object Trajectory Estimation Using Backward Simulation

Yuxuan Xia, Lennart Svensson, Ángel F. García-Fernández, Jason L. Williams, Daniel Svensson, and Karl Granström

Abstract—This paper presents a general solution for computing the multi-object posterior for sets of trajectories from a sequence of multi-object (unlabelled) filtering densities and a multi-object dynamic model. Importantly, the proposed solution opens an avenue of trajectory estimation possibilities for multi-object filters that do not explicitly estimate trajectories. In this paper, we first derive a general multi-trajectory backward smoothing equation based on random finite sets of trajectories. Then we show how to sample sets of trajectories using backward simulation for Poisson multi-Bernoulli filtering densities, and develop a tractable implementation based on ranked assignment. The performance of the resulting multi-trajectory particle smoothers is evaluated in a simulation study, and the results demonstrate that they have superior performance in comparison to several state-of-the-art multi-object filters and smoothers.

Index Terms—Multi-object tracking, random finite sets, sets of trajectories, forward-backward smoothing, backward simulation.

I. INTRODUCTION

Multi-object tracking (MOT) refers to the problem of jointly estimating the number of objects and their trajectories from noisy sensor measurements [1]–[4]. Vector-type MOT methods, e.g., the joint probabilistic data association filter (JPDAF) [5] and the multiple hypothesis tracker (MHT) [6], [7], describe the multi-object states and measurements by random vectors; they explicitly estimate trajectories by linking a state estimate with a previous state estimate or declare the appearance of a new object. However, for MOT methods based on sets representation of the multi-object states, e.g., [8]–[11], sequences of object states at consecutive time steps cannot be easily constructed.

For these MOT methods, one approach to estimating trajectories is to add a unique label to each single-object state such that each object can be identified over time [12]–[16]. This track labelling procedure may work well in some cases, but it often becomes problematic in challenging scenarios, for example, where initially well-separated objects move in close proximity with each other and thereafter separate again

Y. Xia and L. Svensson are with the Department of Electrical Engineering, Chalmers University of Technology, Gothenburg, Sweden. E-mail: first-name.lastname@chalmers.se. A. F. García-Fernández is with the Department of Electrical Engineering and Electronics, University of Liverpool, Liverpool, United Kingdom, and also with the ARIES research centre, Universidad Antonio de Nebrija, Madrid, Spain. J. L. Williams is with the Commonwealth Scientific and Industrial Research Organization, Brisbane, Australia. D. Svensson is with NVIDIA Corporation, Gothenburg, Sweden. K. Granström is with Embark Trucks Inc., San Francisco, CA, USA.

The work of D. Svensson was done when he was with Zenseact AB, and the work of K. Granström was done when he was with Chalmers University of Technology.

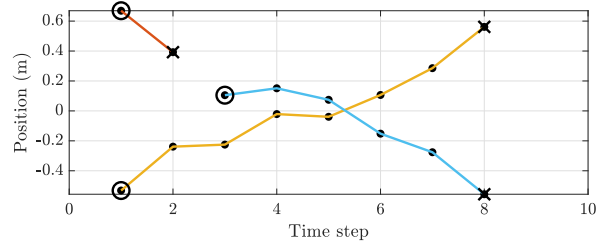


Fig. 1. Illustration of a one-dimensional example where the multi-object filtering density at each time step is a Dirac delta, whose corresponding object states are shown as black dots. The trajectory building problem in this case reduces to how to connect the dots at consecutive time steps. The solid lines represent one of the many ways to construct a set of trajectories where start/end positions of trajectories are marked by circles/crosses. This example will be further elaborated in Section VI-A.

[16]–[18]. A more advantageous approach to estimating trajectories for filters based on random finite sets (RFSs) [8] is to compute the multi-object posterior on sets of trajectories [17], which captures all the information about the trajectories. This has led to the development of a variety of multi-object trackers: the trajectory probability hypothesis density (PHD) filter [19], the trajectory cardinality PHD filter [19], the trajectory Poisson multi-Bernoulli mixture (PMBM) filter [18], the trajectory multi-Bernoulli mixture (MBM) filter [20], and their approximations the trajectory Poisson multi-Bernoulli (PMB) filter [21] and the trajectory multi-Bernoulli (MB) filter [22]. Note that for RFS-based filters with MB birth, the multi-object posterior may be labelled to consider sets of labelled trajectories [17], [20], [22], [23].

Smoothing for state-space models considers the estimation of object states of interest conditioned on the complete measurement sequence [24]. Therefore, smoothing may provide significantly better object state estimation performance than filtering, by refining earlier object state estimates. Solutions to single-object smoothing in clutter mainly include the Gaussian sum smoother [25]–[29], the (integrated) probabilistic data association smoother [30]–[32] and the Bernoulli smoother [33]–[35]. For vector-type MOT methods with Gaussian filtering, smoothed trajectory estimates are typically obtained by applying a Rauch-Tung-Striebel (RTS) smoother [36] on sequences of single-object filtering densities [37]–[39]. It is also possible to consider a batch solution for estimating trajectories using expectation-maximisation [40]. As a comparison, trajectory filters [17]–[23] recursively compute the posterior of sets of trajectories as new observations arrive, by performing smoothing-while-filtering [41], and therefore the initiation and

termination of estimated trajectories can also be improved.

Nevertheless, there are several MOT methods in the literature, e.g., the set JPDAF [42] and the variational PMB filter [43], that can efficiently estimate the multi-object states, but that cannot easily produce smoothed trajectory estimates in a principled manner¹. Then an interesting research question arises: “*Can we leverage filters that do not keep trajectory information to compute the posterior density of sets of trajectories?*”. A one-dimensional example of such trajectory building is illustrated in Fig. 1. Also note that, even for methods that do retain implicit trajectory information, in complex scenarios, approximations made for computational tractability (such as pruning) may cause loss of information that could be recovered with the help of information from later time steps.

In this paper, we show that the exact multi-object posterior of sets of trajectories can be obtained from a sequence of multi-object (unlabelled) filtering densities using the multi-object dynamic model. Specifically, we present a general multi-trajectory backward smoothing equation based on sets of trajectories. The proposed solution has important advantages over multi-object forward-backward smoothers that only compute the marginal multi-object smoothing densities at each time step [45]–[52], which, even if labelled, may not be enough to provide meaningful trajectory information [17, Example 2]. Moreover, the proposed forward-backward smoother does not specify the form of the multi-object filtering densities. This is in contrast to multi-object forward-backward smoothers based on labelled RFSs [53]–[55], which cannot incorporate the Poisson birth model and require that the multi-object filtering densities must be labelled. The outcome of this work is a method for efficiently sampling multi-object trajectories from the posterior distribution of sets of trajectories, based on operations involving only the single time step multi-object state distributions constructed during forward filtering. This has important applications to offline trajectory analytics, e.g., extracting trajectory estimates that can be viewed as ground truth for the development and verification of perception modules in autonomous driving.

A preliminary version of this work was presented in [56]. This paper is a significant extension of that work, and contains the following contributions:

- 1) We derive a general multi-trajectory backward smoothing equation based on sets of trajectories.
- 2) We propose a multi-trajectory particle smoother using backward simulation [57] for PMB filtering densities. In particular, this is a method for doing inference (drawing samples) in the multi-object trajectory space while only ever manipulating single time step, single object marginal distributions, which is only possible because of the properties of the multi-trajectory dynamic model that is stated in Section II-C.
- 3) We present a tractable implementation of the proposed smoother based on a linear-Gaussian dynamic model and ranked assignment.

- 4) We compare the proposed algorithm to several state-of-the-art algorithms [18], [23], [58] in a simulation study, and the results demonstrate that the proposed methods have superior performance.

The rest of the paper is organised as follows. The background on sets of trajectories, multi-object dynamic models and PMB filtering densities is introduced in Section II. The backward smoothing equation for sets of trajectories is presented in Section III. A multi-trajectory particle smoother for PMB filtering densities and its tractable implementation are given in Section IV and Section V, respectively. The simulation results are shown in Section VI, and the conclusions are drawn in Section VII.

II. BACKGROUND

The notation in this paper is defined following the convention in [17], [59]. For a generic space D , the set of finite subsets of D is denoted by $\mathcal{F}(D)$, and the cardinality of a set $A \in \mathcal{F}(D)$ is $|A|$. The sequence of ordered positive integers $(\alpha, \alpha + 1, \dots, \gamma - 1, \gamma)$ is denoted by $\alpha : \gamma$, and the set that includes all the permutations of $1 : n$ is denoted by Γ_n . We use \uplus to denote union of sets that are mutually disjoint, $\langle f, g \rangle$ to denote the inner product $\int f(x)g(x)dx$, and the multi-object exponential f^A , for some real-valued function f , to denote the product $\prod_{x \in A} f(x)$ with $f^\emptyset = 1$ by convention. In addition, we use $\delta_x(\cdot)$ and $\delta_x[\cdot]$ to represent the Dirac and Kronecker delta functions centred at x , respectively.

A. State variables

The single-object state is described by a vector $x \in \mathbb{R}^{n_x}$, typically containing the kinematic information about the object (e.g., position and velocity). A trajectory is represented as a variable $X = (t, x^{1:\nu})$ where t is the initial time step of the trajectory, ν is its length, and $x^{1:\nu} = (x^1, \dots, x^\nu)$ denotes a finite sequence of length ν that contains the object states at time steps $t : t + \nu - 1$. For two time steps α and γ , $\alpha \leq \gamma$, a trajectory $(t, x^{1:\nu})$ in the time interval $\alpha : \gamma$ existing from time step t to $t + \nu - 1$ satisfies that $\alpha \leq t \leq t + \nu - 1 \leq \gamma$, and the variable (t, ν) hence belongs to the set $I_{(\alpha, \gamma)} = \{(t, \nu) : \alpha \leq t \leq \gamma \text{ and } 1 \leq \nu \leq \gamma - t + 1\}$. A single trajectory in the time interval $\alpha : \gamma$ therefore belongs to the space $T_{(\alpha, \gamma)} = \uplus_{(t, \nu) \in I_{(\alpha, \gamma)}} \{t\} \times \mathbb{R}^{\nu n_x}$. We note that trajectory X is a combination of discrete and continuous states. Such a hybrid state is not uncommon in MOT: a typical example is the interacting multiple model [60].

A set $\mathbf{x} \in \mathcal{F}(\mathbb{R}^{n_x})$ of single-object states is a finite subset of \mathbb{R}^{n_x} , and a set $\mathbf{X}_{\alpha:\gamma} \in \mathcal{F}(T_{(\alpha, \gamma)})$ of trajectories is a finite subset of $T_{(\alpha, \gamma)}$. The subset of trajectories in $\mathbf{X}_{\alpha:\gamma}$ that were alive at time step η where $\alpha \leq \eta \leq \gamma$ is denoted by

$$\mathbf{X}_{\alpha:\gamma}^\eta = \{(t, x^{1:\nu}) \in \mathbf{X}_{\alpha:\gamma} : t \leq \eta \leq t + \nu - 1\}.$$

Given a set of trajectories $\mathbf{X}_{\alpha:\gamma}$, we denote the resulting set of trajectories in the time interval $\eta : \zeta$ by

$$\mathbf{X}_{\eta:\zeta} = \left\{ (\epsilon, x^{\epsilon-\eta+1:\epsilon+1}) : (t, x^{1:\nu}) \in \mathbf{X}_{\alpha:\gamma}, \right. \\ \left. \epsilon = \max(\eta, t), \iota = \min(\zeta, t + \nu - 1), \epsilon \leq \iota \right\}.$$

¹A heuristic track-to-target management scheme was presented in [44] by making use of the permutation probabilities of state vectors.

Note that $\mathbf{X}_{\eta;\zeta}$ depends on $\mathbf{X}_{\alpha;\gamma}$, but we keep this relation implicit for notational clarity. An illustrative example of $\mathbf{X}_{\alpha;\gamma}$, $\mathbf{X}_{\alpha;\gamma}^\eta$ and $\mathbf{X}_{\eta;\zeta}$ is given in Fig. 2.

Given a single-object trajectory $X = (t, x^{1:\nu})$, the set of object states at time step k is

$$\tau^k(X) = \begin{cases} \{x^{k+1-t}\}, & t \leq k \leq t + \nu - 1 \\ \emptyset, & \text{otherwise} \end{cases}$$

and given a set $\mathbf{X}_{\alpha;\gamma}$ of trajectories, the set of object states at time step k is $\tau^k(\mathbf{X}_{\alpha;\gamma}) = \bigcup_{X \in \mathbf{X}_{\alpha;\gamma}} \tau^k(X)$.

B. Densities and integrals

Given a real-valued function $\pi(\cdot)$ on the single trajectory space $T_{(\alpha,\gamma)}$, its integral is [17]

$$\int \pi(X) dX = \sum_{(t,\nu) \in I_{(\alpha,\gamma)}} \int \pi(t, x^{1:\nu}) dx^{1:\nu}, \quad (1)$$

which goes through all possible start times, lengths and object states of trajectory $X \in T_{(\alpha,\gamma)}$. The single trajectory space is locally compact, Hausdorff and second-countable [17], and therefore one can perform inference on sets of finite number of trajectories with finite length [8].

Given a real-valued function $\pi(\cdot)$ on the space $\mathcal{F}(T_{(\alpha,\gamma)})$ of finite sets of trajectories, its set integral is [17]

$$\int \pi(\mathbf{X}) \delta \mathbf{X} = \pi(\emptyset) + \sum_{n=1}^{\infty} \frac{1}{n!} \int \pi(\{X_1, \dots, X_n\}) dX_{1:n} \quad (2)$$

where $X_{1:n} = (X_1, \dots, X_n)$. A function $\pi(\cdot)$ on the space $\mathcal{F}(T_{(\alpha,\gamma)})$ is a multi-trajectory density if $\pi(\cdot) \geq 0$ and its set integral is one.

The multi-object state density at time step k' and the multi-trajectory density in the time interval $\alpha : \gamma$, both conditioned on the sequence of sets of measurements $\mathbf{z}_{1:k} = (\mathbf{z}_1, \dots, \mathbf{z}_k)$ received up to and including time step k , are denoted by $f_{k'|k}(\cdot)$ and $\pi_{\alpha;\gamma|k}(\cdot)$, respectively. Given a set \mathbf{x}_k of object states at time step k , the set of trajectories in the time interval $k : k$ is

$$\mathbf{X}_{k:k} = \{X = (k, x^1) : x^1 \in \mathbf{x}_k\}$$

where trajectory $X = (t, x^{1:\nu}) \in \mathbf{X}_{k:k}$ has start time $t = k$ and length $\nu = 1$ with probability one. Therefore, it holds that the multi-trajectory density $\pi_{k':k'|k}(\mathbf{X}_{k':k'})$ takes the same value as the multi-object state density $f_{k'|k}(\mathbf{x})$ when $\tau^{k'}(\mathbf{X}_{k':k'}) = \mathbf{x}$ where $k' \in \{k, k+1\}$.

A trajectory Poisson point process (PPP) has density

$$\pi(\mathbf{X}_{\alpha;\gamma}) = e^{-\langle \lambda, 1 \rangle} [\lambda(\cdot)]^{\mathbf{X}_{\alpha;\gamma}} \quad (3)$$

where $\lambda(\cdot)$ is the Poisson intensity, and a trajectory Bernoulli process has density

$$\pi(\mathbf{X}_{\alpha;\gamma}) = \begin{cases} 1 - r, & \mathbf{X}_{\alpha;\gamma} = \emptyset \\ rp(X), & \mathbf{X}_{\alpha;\gamma} = \{X\} \\ 0, & \text{otherwise} \end{cases} \quad (4)$$

where $p(\cdot)$ is a single-trajectory density and r is the probability of existence. A trajectory MB process is the union of $n \geq 1$

independent trajectory Bernoulli components and its density is given by the convolution formula for RFSs [8, Sec. 11.5.3]

$$\pi(\mathbf{X}_{\alpha;\gamma}) = \sum_{\biguplus_{i=1}^n \mathbf{X}^i = \mathbf{X}_{\alpha;\gamma}} \prod_{i=1}^n \pi^i(\mathbf{X}^i) \quad (5)$$

where $\pi^i(\cdot)$ is the density of the i -th Bernoulli component, and the sum in (5) goes through all disjoint and possibly empty subsets $\mathbf{X}^1, \dots, \mathbf{X}^n$ such that $\mathbf{X}^1 \cup \dots \cup \mathbf{X}^n = \mathbf{X}_{\alpha;\gamma}$.

Given two single-object trajectories $X = (t, x^{1:\nu})$ and $Y = (t', y^{1:\nu'})$, the trajectory Dirac delta function is defined as

$$\delta_Y(X) = \delta_{t'}[t] \delta_{\nu'}[\nu] \delta_{y^{1:\nu'}}(x^{1:\nu}),$$

and the multi-trajectory Dirac delta function centred at \mathbf{Y} is defined as [8, Sec. 11.3.4.3]

$$\delta_{\mathbf{Y}}(\mathbf{X}) = \begin{cases} 0, & |\mathbf{X}| \neq |\mathbf{Y}| \\ 1, & \mathbf{X} = \mathbf{Y} = \emptyset \\ \sum_{\sigma \in \Gamma_n} \prod_{i=1}^n \delta_{Y_{\sigma_i}}(X_i) & \begin{cases} \mathbf{X} = \{X_i\}_{i=1}^n \\ \mathbf{Y} = \{Y_i\}_{i=1}^n \end{cases} \end{cases}$$

C. Multi-trajectory dynamic model

The conventional multi-object dynamic model described in [8] is considered. Given the current multi-object state \mathbf{x} , each object $x \in \mathbf{x}$ survives with probability $p^S(x)$, and moves to a new state with a Markovian transition density $g(\cdot|x)$, or dies with probability $1 - p^S(\cdot)$. The multi-object state at the next time step is the union of the surviving objects and new objects, which are born independently of the rest. The newborn objects are typically modelled as a PPP or an MB process with multi-object state density $\beta(\cdot)$.

The above conventional multi-object dynamic model results in the following dynamic model for the set of all trajectories that have existed up to the current time step, which will be required in backward smoothing for sets of trajectories. Given a set $\mathbf{X}_{1:k}$ of all trajectories in the time interval $1 : k$, each trajectory $X = (t, x^{1:\nu}) \in \mathbf{X}_{1:k}$ ‘‘survives’’ with probability one, $p^S(X) = 1$, and moves to a new state according to [17]

$$g^{k+1}(t', y^{1:\nu'} | X) = |\tau^k(X)| \left[(1 - p^S(x^\nu)) \delta_X(t', y^{1:\nu'}) + p^S(x^\nu) g(y^{\nu'} | x^\nu) \delta_X(t', y^{1:\nu'-1}) \right] + (1 - |\tau^k(X)|) \delta_X(t', y^{1:\nu'}). \quad (6)$$

That is, if the object underlying trajectory X has died before time step k , the trajectory remains unaltered with probability one. If trajectory X exists at time step k , it remains unaltered with probability $1 - p^S(x^\nu)$, or the new final object state $y^{\nu'}$ is generated according to the single-object transition density with probability $p^S(x^\nu)$. The set $\mathbf{X}_{1:k+1}$ of trajectories in the time interval $1 : k+1$ is the union of the dead trajectories, surviving trajectories and new trajectories where each new trajectory $(t, x^{1:\nu})$ has deterministic start time $t = k+1$, length $\nu = 1$, and multi-object state is distributed as $\beta(\cdot)$.

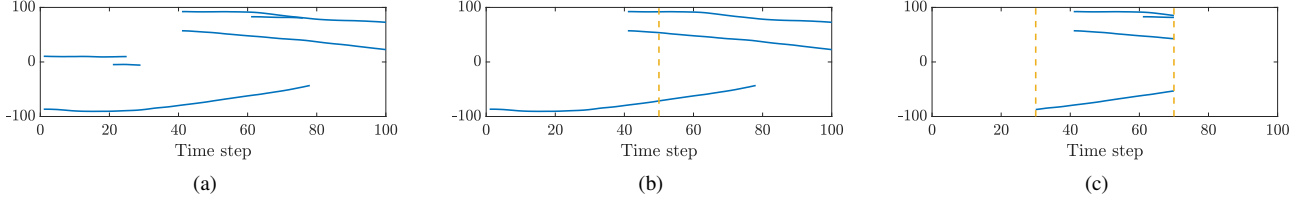


Fig. 2. One-dimensional examples of sets of trajectories: (a) a set $\mathbf{X}_{1:100}$ of six trajectories in the time interval 1 : 100; (b) the set $\mathbf{X}_{1:100}^{50}$ of three trajectories in the time interval 1 : 100 that were alive at time step 50; (c) the resulting set $\mathbf{X}_{30:70}$ of four trajectories in the time interval 30 : 70.

D. PMB filtering densities

RFS-based Bayes filters propagate the multi-object posterior density of \mathbf{x}_k in time via the prediction and update steps:

$$f_{k|k-1}(\mathbf{x}) = \int g(\mathbf{x}|\mathbf{x}') f_{k-1|k-1}(\mathbf{x}') \delta \mathbf{x}', \quad (7)$$

$$f_{k|k}(\mathbf{x}) = \frac{\ell(\mathbf{z}_k|\mathbf{x}) f_{k|k-1}(\mathbf{x})}{\int \ell(\mathbf{z}_k|\mathbf{x}) f_{k|k-1}(\mathbf{x}) \delta \mathbf{x}} \quad (8)$$

where $g(\cdot|\mathbf{x})$ is the multi-object transition density and $\ell(\mathbf{z}_k|\cdot)$ is the measurement likelihood. For the multi-object dynamic model described in Section II-C with Poisson birth model

$$\beta(\mathbf{x}) = e^{-\langle \lambda_k^B, 1 \rangle} [\lambda_k^B(\cdot)]^{\mathbf{x}} \quad (9)$$

where $\lambda_k^B(\cdot)$ is the Poisson birth intensity at time step k , and general multi-object measurement models with Poisson clutter, if the prior is PMBM, the predicted and posterior densities on the current set of object states are PMBM [61]. We note that the Poisson birth model is a special case of PMBM.

The PMB is a common and efficient approximation of a PMBM [11], [43], and its filtering density is defined as

$$f_{k'|k}(\mathbf{x}) = \sum_{\mathbf{x}_d \uplus \mathbf{x}_u = \mathbf{x}} f_{k'|k}^p(\mathbf{x}_u) f_{k'|k}^{mb}(\mathbf{x}_d), \quad (10a)$$

$$f_{k'|k}^p(\mathbf{x}) = e^{-\langle \lambda_{k'}^u, 1 \rangle} [\lambda_{k'}^u(\cdot)]^{\mathbf{x}}, \quad (10b)$$

$$f_{k'|k}^{mb}(\mathbf{x}) = \sum_{\uplus_{i=1}^{n_{k'|k}} \mathbf{x}_d^i = \mathbf{x}} \prod_{i=1}^{n_{k'|k}} f_{k'|k}^i(\mathbf{x}_d^i), \quad (10c)$$

$$f_{k'|k}^i(\mathbf{x}) = \begin{cases} 1 - r_{k'|k}^i, & \mathbf{x} = \emptyset \\ r_{k'|k}^i p_{k'|k}^i(x), & \mathbf{x} = \{x\} \\ 0, & \text{otherwise} \end{cases} \quad (10d)$$

with $k' \in \{k, k+1\}$. The PMB is the union of two independent RFSs: a PPP with density $f_{k'|k}^p(\cdot)$, parameterised by Poisson intensity $\lambda_{k'}^u(\cdot)$, representing undetected objects, and an MB process with density $f_{k'|k}^{mb}(\cdot)$ representing detected objects where the i -th Bernoulli component has density $f_{k'|k}^i(\cdot)$ with probability of existence $r_{k'|k}^i$ and single-object state density $p_{k'|k}^i(\cdot)$.

III. BACKWARD SMOOTHING FOR SETS OF TRAJECTORIES

In this paper, the objective is to compute the multi-trajectory posterior density $\pi_{1:K|K}(\mathbf{X}_{1:K})$ using a sequence of multi-object filtering densities $f_{k|k}(\cdot)$ with $k = 1, \dots, K$ and the multi-trajectory dynamic model via backward smoothing. To

achieve this, we first present the multistep prediction theorem for sets of trajectories, which generalises the general prediction theorem for sets of trajectories [17, Theorem 7] to multistep prediction, along with a resulting corollary that is important for the derivation of the general backward smoothing equation for sets of trajectories.

Theorem 1. Given a multi-trajectory density $\pi_{\alpha:\eta|k}(\mathbf{X}_{\alpha:\eta})$, its $(\gamma - \eta)$ -step predicted multi-trajectory density, with $\alpha \leq \eta < \gamma$, $\eta \geq k$ and $\gamma \geq k + 1$, is given by

$$\begin{aligned} \pi_{\alpha:\gamma|k}(\mathbf{X}_{\alpha:\gamma}) &= \prod_{(t, x^{1:\nu}) \in \mathbf{X}_{\alpha:\gamma}^{\eta}} \left[(1 + p^S(x^\nu) (\delta_{\gamma-t+1}[\nu] - 1)) \right. \\ &\times \left. \prod_{\ell=\eta-t+1}^{\nu-1} g(x^{\ell+1}|x^\ell) p^S(x^\ell) \right] \pi_{\alpha:\eta|k}(\mathbf{X}_{\alpha:\eta}) \pi_{\eta+1:\gamma}(\mathbf{W}) \end{aligned} \quad (11)$$

$$\begin{aligned} \pi_{\eta+1:\gamma}(\mathbf{W}) &= \prod_{(t, x^{1:\nu}) \in \mathbf{W}} \left[(1 + p^S(x^\nu) (\delta_{\gamma-t+1}[\nu] - 1)) \right. \\ &\times \left. \prod_{\ell=1}^{\nu-1} g(x^{\ell+1}|x^\ell) p^S(x^\ell) \right] \prod_{\ell=\eta+1}^{\gamma} \beta(\tau^\ell(\mathbf{W}^\ell)) \end{aligned} \quad (12)$$

where we write $\mathbf{X}_{\alpha:\gamma} = \mathbf{X}_{\alpha:\gamma}^{\eta} \uplus \mathbf{W}$ and $\mathbf{W} = \mathbf{W}^{\eta+1} \uplus \dots \uplus \mathbf{W}^{\gamma}$ denotes the set of trajectories born in the time interval $\eta + 1 : \gamma$, where $\mathbf{W}^\ell = \{(t, x^{1:\nu}) \in \mathbf{W} : t = \ell\}$ is the set of trajectories born at time step ℓ with $\eta + 1 \leq \ell \leq \gamma$.

Theorem 1 is proved in Appendix A, and it describes that the $(\gamma - \eta)$ -step predicted multi-trajectory density of $\pi_{\alpha:\eta|k}(\mathbf{X}_{\alpha:\eta})$ can be evaluated by multiplying the following terms: the multi-trajectory density $\pi_{\alpha:\eta|k}(\mathbf{X}_{\alpha:\eta})$, $1 - p^S(\cdot)$ for trajectories that died in the time interval $\eta + 1 : \gamma$, $g(\cdot|\cdot)p^S(\cdot)$ for trajectories that were alive at time step η , and the multi-trajectory density $\pi_{\eta+1:\gamma}(\mathbf{W})$ for trajectories that appeared after time step η .

Corollary 1.1. For $\gamma \geq k + 1$, it holds that

$$\frac{\pi_{k:\gamma|k}(\mathbf{X}_{k:\gamma})}{\pi_{k+1:\gamma|k}(\mathbf{X}_{k+1:\gamma})} = \frac{\pi_{k:k+1|k}(\mathbf{X}_{k:k+1})}{f_{k+1|k}(\tau^{k+1}(\mathbf{X}_{k+1:k+1}))}. \quad (13)$$

Corollary 1.1 is proved in Appendix B, and it shows that the ratio between the two multi-trajectory densities $\pi_{k:\gamma|k}(\mathbf{X}_{k:\gamma})$ and $\pi_{k+1:\gamma|k}(\mathbf{X}_{k+1:\gamma})$ does not depend on γ for $\gamma \geq k + 1$.

The general backward smoothing equation for sets of trajectories under the conventional multi-object dynamic model assumptions is presented in the following theorem. Its proof is given in Appendix C.

Theorem 2. Given the multi-trajectory density $\pi_{k+1:K|K}(\cdot)$ and the multi-object state filtering density $f_{k|k}(\cdot)$, the multi-trajectory density in the time interval $k : K$ conditioned on the sequence of sets of measurements up to and including time step K is

$$\pi_{k:K|K}(\mathbf{X}_{k:K}) = \frac{\pi_{k:k+1|k}(\mathbf{X}_{k:k+1})\pi_{k+1:K|K}(\mathbf{X}_{k+1:K})}{f_{k+1|k}(\tau^{k+1}(\mathbf{X}_{k+1:k+1}))} \quad (14)$$

where the multi-object state predicted density $f_{k+1|k}(\cdot)$ can be computed using $f_{k|k}(\cdot)$ via (7), and $\pi_{k:k+1|k}(\mathbf{X}_{k:k+1})$ is the one-step predicted multi-trajectory density of $\pi_{k:k|k}(\mathbf{X}_{k:k})$, which takes the same value as $f_{k|k}(\tau^k(\mathbf{X}_{k:k}))$.

Theorem 2 shows that the multi-trajectory smoothing density $\pi_{k:K|K}(\mathbf{X}_{k:K})$ can be expressed as the product of the multi-trajectory smoothing density $\pi_{k+1:K|K}(\mathbf{X}_{k+1:K})$ and the one-step predicted multi-trajectory density of $f_{k|k}(\cdot)$, normalised by the multi-object state predicted density $f_{k+1|k}(\cdot)$. By applying Theorem 2 recursively backwards in time, we obtain $\pi_{1:K|K}(\mathbf{X}_{1:K})$, as desired.

IV. A MULTI-TRAJECTORY PARTICLE SMOOTHER

In this section, we first introduce a general backward kernel for sets of trajectories, which enables us to perform backward simulation and to approximate the multi-trajectory posterior. Then we present how to evaluate the backward kernel when the multi-object filtering densities are PMB. For the proposed multi-trajectory particle smoother using PMB filtering densities, the forward filtering representations involving single time step, single object marginal distributions can implicitly represent uncertainty over an exponentially large hypothesis space that cannot be feasibly represented through multiple hypothesis methods. The backward simulation framework allows us to draw samples in the multi-object trajectory space from these single time step, single object marginal distributions calculated during the forward filtering which incorporate information from measurements over all the time steps.

A. Backward kernel for sets of trajectories

The backward kernel for sets of trajectories is introduced in the following lemma. Its proof is given in Appendix D.

Lemma 3. The backward kernel for sets of trajectories, i.e., the multi-trajectory density of the set \mathbf{X} of trajectories in the time interval $k : K$ conditioned on the set \mathbf{Y} of trajectories in the time interval $k + 1 : K$ and the sequence of measurement sets up to and including time step K , satisfies

$$\pi_{k:K|K}(\mathbf{X}|\mathbf{Y}) \propto \pi_{k:k+1|k}(\mathbf{X}_{k:k+1})\delta_{\mathbf{Y}}(\mathbf{X}_{k+1:K}). \quad (15)$$

Lemma 3 describes that the backward kernel $\pi_{k:K|K}(\mathbf{X}|\mathbf{Y})$ conditioned on the set \mathbf{Y} of trajectories is proportional to the one-step predicted multi-trajectory density $\pi_{k:k+1|k}(\mathbf{X}_{k:k+1})$ only if $\mathbf{Y} = \mathbf{X}_{k+1:K}$, and zero otherwise. Note that although the set $\mathbf{X}_{k+1:K}$ of trajectories is deterministically given by \mathbf{Y} , evaluating the backward kernel $\pi_{k:K|K}(\mathbf{X}|\mathbf{Y})$ is complicated by the multiple different ways of associating $\mathbf{X}_{k:k+1}$ to \mathbf{Y} ; see, e.g., Fig. 1.

In both Theorem 2 and Lemma 3, we consider the distribution of the set of trajectories in the time interval $k : K$, and we condition on all measurements in the time interval $1 : K$. The main difference between Theorem 2 and Lemma 3 is that the multi-trajectory density computed in Lemma 3 is conditioned on a particular value of the set \mathbf{Y} of trajectories in the time interval $k + 1 : K$, which is what we require for backward simulation.

B. Backward kernel for PMB filtering densities

We proceed to present the backward kernel (15) for PMB filtering densities, see (10). We first observe that the backward kernel (15) is in analogy to the Bayesian measurement update (8) in the sense that $\pi_{k:k+1|k}(\mathbf{X}_{k:k+1})$ is the prior and $\delta_{\mathbf{Y}}(\mathbf{X}_{k+1:K})$ is the measurement likelihood. Specifically, for a PMB filtering density $f_{k|k}(\cdot)$, the one-step predicted multi-trajectory density $\pi_{k:k+1|k}(\mathbf{X}_{k:k+1})$ is a trajectory PMB [21, Lemma 4]. Moreover, the multi-trajectory Dirac delta $\delta_{\mathbf{Y}}(\mathbf{X}_{k+1:K})$ can be understood as a standard multi-object measurement model [8] with the following characteristics (this will be further elaborated in Appendix E):

- Each trajectory $X = (t, x^{1:\nu}) \in \mathbf{X}_{k:K}$ is detected with probability

$$p^D(X) = \begin{cases} 0, & t = k \text{ and } \nu = 1 \\ 1, & \text{otherwise} \end{cases} \quad (16)$$

and if detected, it generates a measurement Y with density $\delta_Y(X_{k+1:K})$.

- Clutter is Poisson with intensity $\lambda^C(\cdot) = 0$.

Then, the backward kernel $\pi_{k:K|K}(\mathbf{X}|\mathbf{Y})$ (15) for PMB filtering densities, obtained via updating the trajectory PMB prior $\pi_{k:k+1|k}(\mathbf{X}_{k:k+1})$ using $\delta_{\mathbf{Y}}(\mathbf{X}_{k+1:K})$ is a trajectory PMBM [21, Lemma 5]. In what follows, we explain the result of the trajectory PMB update applied to backward simulation with PMB filtering densities, where each trajectory Bernoulli component has information on start/end times of the trajectories giving rise to the following local hypotheses:

- The trajectory had never existed in the entire time interval $1 : k + 1$.
- The trajectory ended at time step k .
- The trajectory existed at both time step k and $k + 1$.
- The trajectory started at time $k + 1$.

We define by $\mathcal{M}_{k+1:K} = \{1, \dots, n_{k+1:K}\}$ the set of indices of trajectories in \mathbf{Y} . Each trajectory in \mathbf{Y} creates a unique trajectory Bernoulli component. The number of trajectory Bernoulli components in $\pi_{k:K|K}(\mathbf{X}|\mathbf{Y})$ is $n_{k:K|K} = n_{k|k} + n_{k+1:K}$, and they are indexed by variable $i \in \{1, \dots, n_{k:K|K}\}$. A global hypothesis is $a = (a^1, \dots, a^{n_{k:K|K}})$, where h^i is the number of local hypotheses and $a^i \in \{1, \dots, h_{k:K|K}^i\}$ is the index to the local hypothesis for the i -th trajectory Bernoulli component, which incorporates the following information:

- The set of indices $\mathcal{M}_{k+1:K}^{i,a^i} \subseteq \mathcal{M}_{k+1:K}$. If $\mathcal{M}_{k+1:K}^{i,a^i} = \emptyset$, then the i -th hypothesised trajectory did not appear after time step k . If $\mathcal{M}_{k+1:K}^{i,a^i} = \{j\}$, then the i -th hypothesised trajectory existed in the time interval $k + 1 : K$.

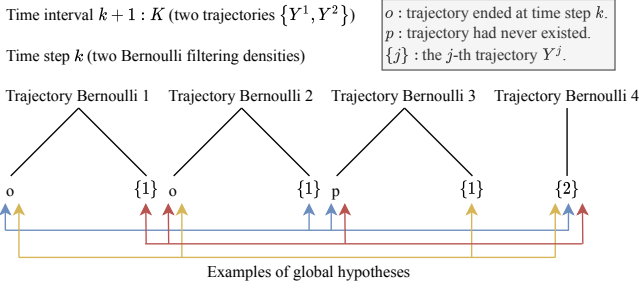


Fig. 3. Hypothesis structure of the PMBM backward kernel for the example illustrated above where there are two trajectories in the time interval $k+1 : K$ and two Bernoulli filtering densities at time step k . For the two trajectories: Y^1 is the trajectory of an object that existed at time step $k+1$, and Y^2 is the trajectory of an object that appeared after time step $k+1$. There are four trajectory Bernoulli components. For each trajectory Bernoulli component in $\pi_{k:k+1|k}(\mathbf{X}_{k:k+1})$ (Trajectory Bernoulli 1 and 2), each has two local hypotheses: one corresponds to the case that the trajectory ended at time step k , and the other corresponds to the case that the trajectory Bernoulli component was updated by trajectory Y^1 . The trajectory Bernoulli component created by trajectory Y^1 (Trajectory Bernoulli 3) has two local hypotheses: one corresponds to the case that the trajectory started at time step $k+1$, and the other corresponds to the case that the trajectory never existed. The trajectory Bernoulli component created by trajectory Y^2 (Trajectory 4) has only a single local hypothesis since Y^2 remains unaltered.

- The local hypothesis weight $w_{k:k|K}^{i,a^i}$ (used for computing the probability of the global hypotheses).
- The hypothesis-conditioned Bernoulli density $\pi_{k:k|K}^{i,a^i}(\cdot)$ of the form (4), parameterised by the probability of existence $r_{k:k|K}^{i,a^i}$ and the single-trajectory density $p_{k:k|K}^{i,a^i}(\cdot)$.

The set of all global hypotheses is

$$\mathcal{A}_{k:k|K} = \left\{ (a^1, \dots, a^{n_{k:k|K}}) : a^i \in \{1, \dots, h_{k:k|K}^i\} \forall i, \left| \mathcal{M}_k^{i,a^i} \right| \leq 1, \bigoplus_{i=1}^{n_{k:k|K}} \mathcal{M}_k^{i,a^i} = \mathcal{M}_k \right\}, \quad (17)$$

and the weight of global hypothesis $a \in \mathcal{A}_{k:k|K}$ satisfies

$$w^a \propto \prod_{i=1}^{n_{k:k|K}} w_{k:k|K}^{i,a^i} \quad (18)$$

where the proportionality means that normalisation is required to ensure that $\sum_{a \in \mathcal{A}_{k:k|K}} w^a = 1$. A simple example of the hypothesis structure is illustrated and described in Fig. 3.

The following theorem gives the explicit expression of the PMBM backward kernel (15) for PMB filtering densities.

Theorem 4. *Given a PMB filtering density $f_{k|k}(\cdot)$ (10) at time step k , the set \mathbf{Y} of trajectories in the time interval $k+1 : K$ and the multi-trajectory dynamic model described in Section II-C with Poisson birth (9), the multi-trajectory density in the time interval $k : K$ conditioned on \mathbf{Y} and the sequence of measurement sets up to and including time step K , is a PMBM of the form*

$$\pi_{k:k|K}(\mathbf{X}|\mathbf{Y}) = \sum_{\mathbf{X}_u \uplus \mathbf{X}_d = \mathbf{X}} \pi_{k:k|K}^u(\mathbf{X}_u) \pi_{k:k|K}^d(\mathbf{X}_d|\mathbf{Y}), \quad (19)$$

$$\pi_{k:k|K}^u(\mathbf{X}_u) = e^{-\langle \lambda_{k:k|K}^u, \mathbf{1} \rangle} \left[\lambda_{k:k|K}^u(\cdot) \right]^{\mathbf{X}_u}, \quad (20)$$

$$\pi_{k:k|K}^d(\mathbf{X}_d|\mathbf{Y}) = \sum_{a \in \mathcal{A}_{k:k|K}} w^a \sum_{\biguplus_{j=1}^{n_{k:k|K}} \mathbf{X}^j = \mathbf{X}_d} \prod_{i=1}^{n_{k:k|K}} \pi_{k:k|K}^{i,a^i}(\mathbf{X}^i) \quad (21)$$

where $n_{k:k|K} = n_{k+1:K} + n_{k|k}$, w^a is given by (18) and

$$\lambda_{k:k|K}^u(t, x^{1:\nu}) = \delta_k[t] \delta_1[\nu] (1 - p^S(x^1)) \lambda_{k|k}^u(x^1). \quad (22)$$

Also, we write $\mathbf{Y} = \{Y^1, \dots, Y^{n_{k+1:K}}\}$ with Y^1, \dots, Y^m being trajectories of objects that existed at time step $k+1$, and $Y^{m+1}, \dots, Y^{n_{k+1:K}}$ being trajectories of objects that appeared after time step $k+1$.

For each trajectory Bernoulli component in the predicted trajectory PMB $\pi_{k:k+1|k}(\mathbf{X}_{k:k+1})$, $i \in \{1, \dots, n_{k|k}\}$, there are $h_{k:k|K}^i = m+1$ local hypotheses. The local hypothesis, corresponding to the case that the trajectory ended at time step k , is given by $\mathcal{M}_{k+1:K}^{i,1} = \emptyset$ and

$$w_{k:k|K}^{i,1} = 1 - r_{k|k}^i + r_{k|k}^i \langle p_{k|k}^i, 1 - p^S \rangle, \quad (23a)$$

$$r_{k:k|K}^{i,1} = \frac{r_{k|k}^i \langle p_{k|k}^i, 1 - p^S \rangle}{w_{k:k|K}^{i,1}}, \quad (23b)$$

$$p_{k:k|K}^{i,1}(t, x^{1:\nu}) = \delta_k[t] \delta_1[\nu] \frac{p_{k|k}^i(x^1) (1 - p^S(x^1))}{\langle p_{k|k}^i, 1 - p^S \rangle}. \quad (23c)$$

The local hypothesis, corresponding to the case that the trajectory Bernoulli component is updated by trajectory $Y^j = (t^j, y^{1:\nu^j})$, $j \in \{1, \dots, m\}$ (present at time step $k+1$, i.e., $t^j = k+1$), is given by $\mathcal{M}_{k+1:K}^{i,j+1} = \{j\}$ and

$$w_{k:k|K}^{i,j+1} = r_{k|k}^i \langle p_{k|k}^i, g(y^1|\cdot) p^S \rangle, \quad (24a)$$

$$r_{k:k|K}^{i,j+1} = 1, \quad (24b)$$

$$p_{k:k|K}^{i,j+1}(t, x^{1:\nu}) = \delta_k[t] \delta_{\nu^j+1}[\nu] \delta_{y^{1:\nu^j}}(x^{2:\nu}) \times \frac{g(y^1|x^1) p_{k|k}^i(x^1) p^S(x^1)}{\langle p_{k|k}^i, g(y^1|\cdot) p^S \rangle}. \quad (24c)$$

The trajectory Bernoulli component created by trajectory $Y^j = (t^j, y^{1:\nu^j})$, $j \in \{1, \dots, m\}$, has two local hypotheses $h_{k:k|K}^i = 2$. The first one corresponds to a non-existent Bernoulli and is given by $\mathcal{M}_{k+1:K}^{i,1} = \emptyset$ and

$$w_{k:k|K}^{i,1} = 1, \quad r_{k:k|K}^{i,1} = 0. \quad (25)$$

The second one is given by $\mathcal{M}_{k+1:K}^{i,2} = \{j\}$ and

$$w_{k:K|K}^{i,2} = \lambda_{k+1}^B(y^1) + \left\langle \lambda_{k|k}^u, g(y^1|\cdot) p^S \right\rangle, \quad (26a)$$

$$r_{k:K|K}^{i,2} = 1, \quad (26b)$$

$$\begin{aligned} p_{k:K|K}^{i,2}(t, x^{1:\nu}) &= \underline{w}_{k:K|K}^{i,2} \delta_{(t^j, y^{1:\nu^j})} (t, x^{1:\nu}) \\ &\quad + \bar{w}_{k:K|K}^{i,2} \delta_k[t] \delta_{\nu^j+1}[\nu] \delta_{y^{1:\nu^j}}(x^{2:\nu}) \\ &\quad \times \frac{g(y^1|x^1) \lambda_{k|k}^u(x^1) p^S(x^1)}{\left\langle \lambda_{k|k}^u, g(y^1|\cdot) p^S \right\rangle}, \end{aligned} \quad (26c)$$

$$\underline{w}_{k:K|K}^{i,2} = \frac{\lambda_{k+1}^B(y^1)}{w_{k:K|K}^{i,2}}, \quad (26d)$$

$$\bar{w}_{k:K|K}^{i,2} = 1 - \underline{w}_{k:K|K}^{i,2}. \quad (26e)$$

The trajectory Bernoulli component created by trajectory Y^j , $j \in \{m+1, \dots, n_{k=1:K}\}$ (not present at time step $k+1$) only has a single local hypothesis $h_{k:K|K}^i = 1$, given by $\mathcal{M}_{k+1:K}^{i,1} = \{j\}$ and

$$w_{k:K|K}^{i,1} = 1, \quad (27a)$$

$$r_{k:K|K}^{i,1} = 1, \quad (27b)$$

$$p_{k:K|K}^{i,1}(X) = \delta_{Y^j}(X). \quad (27c)$$

Proof. See Appendix E. \square

The local hypothesis parameterised by (25) corresponds to a non-existent Bernoulli component. For a global hypothesis that includes this local hypothesis, trajectory Y^j has been assigned to a Bernoulli filtering density. The local hypothesis parameterised by (26) covers the case that the trajectory is associated to the trajectory PPP in $\pi_{k:k+1|k}(\mathbf{X}_{k:k+1})$; i.e., the corresponding object was first detected at time step $k+1$. Note that, for an object that was first detected at time step $k+1$, it might be born at any time before, or at, time step $k+1$. Therefore, there is uncertainty on the start time of its corresponding trajectory.

C. Backward simulation for sets of trajectories

A particle approximation of the multi-trajectory density is

$$\pi(\mathbf{X}_{\alpha}; \gamma) \approx \sum_{i=1}^T w^{(i)} \delta_{\mathbf{X}^{(i)}}(\mathbf{X}_{\alpha}; \gamma) \quad (28)$$

where T is the number of particles and the i -th particle has state $\mathbf{X}^{(i)}$ and weight $w^{(i)}$. By running the backward simulation for sets of trajectories T times for $k = K-1, \dots, 1$ where we recursively draw samples of $\mathbf{X}_{k:K}$ from the backward kernel (15), we can obtain T particles $\{\mathbf{X}_{1:K}^{(i)}\}_{i=1}^T$ representing the multi-trajectory density $\pi_{1:K|K}(\mathbf{X}_{1:K})$ with uniform weights $w^{(i)} = 1/T$. Note that, multiple backward set of trajectories can be generated independently, without having to rerun the forward multi-object filter. That is the complexity of backward simulation for sets of trajectories is linear with the number of particles.

For the PMBM backward kernel (19), it is sufficient to sample sets of trajectories described by the MBM (21), which are trajectories of objects that are hypothesised to have been

detected at some time, possibly also including object states before their first detection. In other words, we do not sample trajectories of objects that are hypothesised to exist but were never detected. To do so, we first sample a global hypothesis to obtain a trajectory MB, and then we sample a single trajectory state for each Bernoulli component. It is, however, generally intractable to exhaustively enumerate all the global hypotheses. Therefore, approximations are needed for a tractable implementation. One such implementation will be presented for a linear-Gaussian dynamic model in the next section.

V. A TRACTABLE IMPLEMENTATION FOR LINEAR-GAUSSIAN DYNAMIC MODEL

In this section, we present a tractable implementation of the proposed multi-trajectory particle smoother for PMB filtering densities with the following assumptions:

Assumption 1. The linear Gaussian dynamic model and the PMB filtering densities are defined as follows

- The survival probabilities are constant, i.e., $p^S(\cdot) = p^S$.
- The linear Gaussian single-object state transition density is $g(\cdot|x) = \mathcal{N}(\cdot; Fx, Q)$ where F is the state transition matrix and Q is the motion noise covariance matrix.
- The Poisson birth intensity is a Gaussian mixture

$$\lambda_k^B(x) = \sum_{i=1}^{N_k^b} w_k^{b,i} \mathcal{N}(x; x_k^{b,i}, P_k^{b,i}) \quad (29)$$

where N_k^b is the number of components, $w_k^{b,i}$, $x_k^{b,i}$ and $P_k^{b,i}$ are the weight, the mean and the covariance of the i -th component, respectively.

- The PMB filtering density at time step k is parameterised by $\left\{ \lambda_{k|k}^u(\cdot), \left\{ r_{k|k}^i, P_{k|k}^i(\cdot) \right\}_{i=1}^{n_{k|k}} \right\}$. Here,

$$\lambda_{k|k}^u(x) = \sum_{i=1}^{N_{k|k}^u} w_{k|k}^{u,i} \mathcal{N}(x; x_{k|k}^{u,i}, P_{k|k}^{u,i}) \quad (30)$$

is the Poisson RFS intensity for undetected objects where $N_{k|k}^u$ is the number of components, $w_{k|k}^{u,i}$, $x_{k|k}^{u,i}$ and $P_{k|k}^{u,i}$ are the weight, the mean and the covariance of the i -th component, respectively. In addition, there are $n_{k|k}$ Bernoulli components, and the i -th component has probability of existence $r_{k|k}^i$ and single-object state density $p_{k|k}^i(x) = \mathcal{N}(x; x_{k|k}^i, P_{k|k}^i)$ where $x_{k|k}^i$ is its mean and $P_{k|k}^i$ is its covariance.

A. Gaussian implementation for backward kernel

The backward kernel for sets of trajectories under Assumption 1 is given by the following lemma. For general non-linear dynamic models, the smoothed state density can be computed using Gaussian assumed density approximations [24].

Lemma 5. Given the linear Gaussian dynamic model, the PMB filtering density at time step k specified in Assumption 1 and the set $\mathbf{Y} = \{Y^1, \dots, Y^{n_{k+1:K}}\}$ of trajectories in the time interval $k+1:K$ specified in Theorem 4, the multi-trajectory

density in the time interval $k : K$ is a PMBM of the form (19) with $n_{k:K|K} = n_{k+1:K} + n_{k|k}$ and Poisson intensity

$$\lambda_{k:K|K}^u(t, x^{1:\nu}) = \delta_k[t] \delta_1[\nu] (1 - p^S) \lambda_{k|k}^u(x^1). \quad (31)$$

The local hypothesis of the i -th trajectory Bernoulli component, $i \in \{1, \dots, n_{k|k}\}$, corresponding to the case that the trajectory ended at time step k , is given by

$$w_{k:K|K}^{i,1} = 1 - r_{k|k}^i + r_{k|k}^i (1 - p^S), \quad (32a)$$

$$r_{k:K|K}^{i,1} = \frac{r_{k|k}^i (1 - p^S)}{w_{k:K|K}^{i,1}}, \quad (32b)$$

$$p_{k:K|K}^{i,1}(t, x^{1:\nu}) = \delta_k[t] \delta_1[\nu] \mathcal{N}(x^1; x_{k|k}^i, P_{k|k}^i). \quad (32c)$$

The local hypothesis of the i -th trajectory Bernoulli component, $i \in \{1, \dots, n_{k|k}\}$, corresponding to the case that the trajectory Bernoulli component is updated by trajectory $Y^j = (t^j, y^{1:\nu^j})$, $j \in \{1, \dots, m\}$, is given by

$$w_{k:K|K}^{i,j+1} = r_{k|k}^i p^S \mathcal{N}(y^1; F x_{k|k}^i, P_{k+1|k}^i), \quad (33a)$$

$$r_{k:K|K}^{i,j+1} = 1, \quad (33b)$$

$$p_{k:K|K}^{i,j+1}(t, x^{1:\nu}) = \delta_k[t] \delta_{\nu^j+1}[\nu] \delta_{y^{1:\nu^j}}(x^{2:\nu}) \times \mathcal{N}(x^1; x_{k|k}^i, P_{k|k}^i), \quad (33c)$$

$$x_{k|k}^i = x_{k|k}^i + G^i (y^1 - F x_{k|k}^i), \quad (33d)$$

$$P_{k|k}^i = P_{k|k}^i - G^i F P_{k|k}^i, \quad (33e)$$

$$G^i = P_{k|k}^i F^T (P_{k+1|k}^i)^{-1}, \quad (33f)$$

$$P_{k+1|k}^i = F P_{k|k}^i F^T + Q. \quad (33g)$$

The local hypothesis of the i -th trajectory Bernoulli component, $i \in \{n_{k|k} + 1, \dots, n_{k|k} + m\}$, corresponding to the case that the object with trajectory $Y^j = (t^j, y^{1:\nu^j})$ was first detected at time step $k + 1$, is given by

$$w_{k:K|K}^{i,2} = \sum_{i=1}^{N_{k+1}^b} w_{k+1}^{b,i} \mathcal{N}(y^1; x_{k+1}^{b,i}, P_{k+1}^{b,i}) + p^S \sum_{i=1}^{N_{k|k}^u} w_{k|k}^{u,i} \mathcal{N}(y^1; F x_{k|k}^{u,i}, P_{k+1|k}^{u,i}), \quad (34a)$$

$$r_{k:K|K}^{i,2} = 1, \quad (34b)$$

$$p_{k:K|K}^{i,2}(t, x^{1:\nu}) = \underline{w}_{k:K|K}^{i,2} \delta_{(t^j, y^{1:\nu^j})}(t, x^{1:\nu}) + \overline{w}_{k:K|K}^{i,2} \delta_k[t] \delta_{\nu^j+1}[\nu] \delta_{y^{1:\nu^j}}(x^{2:\nu}) \times \frac{\sum_{i=1}^{N_{k|k}^u} w_{k|k}^{u,i} \mathcal{N}(x^1; x_{k|k}^{u,i}, P_{k|k}^{u,i})}{\sum_{i=1}^{N_{k|k}^u} w_{k|k}^{u,i}}, \quad (34c)$$

$$x_{k|k}^{u,i} = x_{k|k}^{u,i} + G^i (y^1 - F x_{k|k}^{u,i}), \quad (34d)$$

$$P_{k|k}^{u,i} = P_{k|k}^{u,i} - G^i F P_{k|k}^{u,i}, \quad (34e)$$

$$G^i = P_{k|k}^{u,i} F^T (P_{k+1|k}^{u,i})^{-1}, \quad (34f)$$

$$P_{k+1|k}^{u,i} = F P_{k|k}^{u,i} F^T + Q, \quad (34g)$$

$$\underline{w}_{k:K|K}^{i,2} = \frac{\sum_{i=1}^{N_{k+1}^b} w_{k+1}^{b,i} \mathcal{N}(y^1; x_{k+1}^{b,i}, P_{k+1}^{b,i})}{w_{k:K|K}^{i,2}}, \quad (34h)$$

$$\overline{w}_{k:K|K}^{i,2} = 1 - \underline{w}_{k:K|K}^{i,2}. \quad (34i)$$

The local hypothesis of the i -th trajectory Bernoulli component, $i \in \{n_{k|k} + j\}$, $j \in \{m + 1, \dots, n_{k+1:K}\}$, corresponding to the case that trajectory Y^j remains unaltered, has $w_{k:K|K}^{i,1} = 1$, $r_{k:K|K}^{i,1} = 1$ and $p_{k:K|K}^{i,1}(X) = \delta_{Y^j}(X)$.

B. Practical considerations

The difficulty in drawing samples of sets of trajectories from the trajectory PMBM backward kernel (19) is the number of global hypotheses since enumerating every global hypothesis is of combinatorial complexity. A simple solution is to truncate the MB mixture (21) to only keep global hypotheses with non-negligible weights. This can be achieved by solving a ranked assignment problem using, e.g., Murty's algorithm [62].

As trajectory Bernoulli components created by trajectories $\{Y^j\}_{j=m+1}^{n_{k+1:K}}$ not present at time step $k + 1$ all have a single local hypothesis with weight $w_{k:K|K}^{i,1} = 1$, $i \in \{n_{k|k} + m + 1, \dots, n_{k:K|K}\}$, the global hypothesis weight (18) becomes

$$w^a \propto \prod_{i=1}^{n_{k|k}} w_{k:K|K}^{i,a^i} \prod_{i=n_{k|k}+1}^{n_{k|k}+m} w_{k:K|K}^{i,a^i} \propto \prod_{i=1: a^i > 1}^{n_{k|k}} \frac{w_{k:K|K}^{i,a^i}}{w_{k:K|K}^{i,1}} \prod_{i=n_{k|k}+1}^{n_{k|k}+m} w_{k:K|K}^{i,a^i}. \quad (35)$$

We can then construct the corresponding cost matrix as:

$$C = -\log [W_1 \quad W_2], \quad (36a)$$

$$W_1^{(j,i)} = \frac{w_{k:K|K}^{i,j+1}}{w_{k:K|K}^{i,1}}, \quad (36b)$$

$$W_2 = \text{diag} \left(w_{k:K|K}^{n_{k|k}+1,2}, \dots, w_{k:K|K}^{n_{k|k}+m,2} \right) \quad (36c)$$

where $W_2 \in \mathbb{R}^{m \times m}$ is a diagonal matrix, $W_1 \in \mathbb{R}^{m \times n_{k|k}}$ and its (j, i) -th entry is $W_1^{(j,i)}$, the weight of associating the j -th trajectory to the i -th trajectory Bernoulli component. Each global hypothesis can be represented as an $m \times (n_{k|k} + m)$ assignment matrix S consisting of 0 or 1 entries such that each row sums to one and that each column sums to zero or one. Then we obtain the M -best global hypotheses that minimise $\text{tr}(S^T C)$ using Murty's algorithm. Note that the assignment problem formulated here is similar to the assignment problem involved in the PMBM filter update where measurements are associated to Bernoulli/PPP components.

In addition to pruning global hypotheses with small weights, we apply ellipsoidal gating to remove unlikely local hypotheses. Specifically, if the squared Mahalanobis distance between

the first state y^1 of trajectory Y^j , $1 \leq j \leq m$ and the predicted density of $\mathcal{N}(x; x_{k|k}^i, P_{k|k}^i)$, i.e.,

$$\text{SMD}(y^1; x_{k|k}^i, P_{k|k}^i) \triangleq \left(y^1 - F x_{k|k}^i \right)^T \left(F P_{k|k}^i F^T + Q \right)^{-1} \left(y^1 - F x_{k|k}^i \right)$$

is larger than a pre-defined threshold Γ_g , we then set $W_1^{(j,i)} = 0$. This also applies to the Gaussian components of (29) and (30). In addition, clustering can be used to further simplify the computations of the assignment problem.

Having discussed strategies on how to efficiently sample the global hypotheses, we proceed to describe how to simplify the sampling of trajectories from the trajectory Bernoulli densities. We note that a fairly large number of particles may be needed to find the mode of the multi-trajectory density $\pi_{1:K|K}(\mathbf{X})$ when the Gaussian covariance matrices in (34c) and (33c) are large. A simple heuristic that helps to quickly find the mode in some situations is to approximate each of these Gaussian densities as a Dirac delta centred at its mean. For instance, in this case the single-trajectory density (33c) becomes

$$\hat{p}_{k:K|K}^{i,j+1}(t, x^{1:\nu}) = \delta_k[t] \delta_{\nu j+1}[\nu] \delta_{x_{k|K}^i}(x^1) \delta_{y^{1:\nu j}}(x^{2:\nu}),$$

and this also avoids the need for computing the corresponding Gaussian covariance $P_{k|K}^i$ (33e).

The robustness of the proposed multi-trajectory particle smoother mainly depends on the number T of particles and the maximum number M of global hypotheses used in backward simulation. For scenarios with large motion noises or PMB filtering densities with many Bernoulli components, we may need large T and M to obtain good smoothing performance. In fact, if the computational complexity is not a concern in offline applications (which is usually the case), one can always use larger T and M to obtain improved results.

Finally, the pseudocode of linear/Gaussian backward simulation for sets of trajectories with PMB filtering densities can be found in Appendix G.

VI. ILLUSTRATIVE EXAMPLE AND SIMULATION RESULTS

In this section, we first illustrate how to perform backward smoothing for sets of trajectories using Theorem 2 via a one-dimensional example shown in Fig. 4a. Then, we evaluate the performance of the proposed multi-trajectory particle smoother for PMB filtering densities in two different two-dimensional scenarios.

A. Illustrative example

We consider a one-dimensional scenario on the line segment $[-1 \text{ m}, 1 \text{ m}]$, of length $K = 8$ time steps, where two initially well-separated objects first approach each other (on the real line), then pause for 1 s, before crossing each other. In the model, objects survive with probability $p^S = 0.95$ and new objects appear according to a Poisson birth model with intensity $\lambda^B(x) = 0.1 \times \text{Uniform}[-1, 1]$. A nearly constant velocity motion model is used with sampling period $T_s = 1$ s and single-object state transition density parameterised by

$$F = \begin{bmatrix} 1 & T_s \\ 0 & 1 \end{bmatrix}, \quad Q = \frac{1}{900} \begin{bmatrix} T_s^3/3 & T_s^2/2 \\ T_s^2/2 & T_s \end{bmatrix}$$

We also assume that the multi-object filtering density at each time step is a multi-object Dirac delta, which is a special type of PMB with zero PPP intensity and each Bernoulli component having probability of existence one and single target density being a Dirac delta. The ground truth and filtering estimates are illustrated in Fig. 4a.

The objective is to compute the multi-trajectory posterior $\pi_{1:K|K}(\mathbf{X}_{1:K})$, which, in this example, can be directly derived using Theorem 2. However, due to the multiple possible associations of single-object Dirac deltas at different time steps, the number of mixture components in the multi-trajectory smoothing density $\pi_{k:K|K}(\mathbf{X}_{k:K})$ increases very fast when computed backward in time, and thus it soon becomes infeasible to evaluate $\pi_{k:K|K}(\mathbf{X}_{k:K})$ without approximation. As shown in Appendix H, the multi-trajectory smoothing density $\pi_{K-1:K|K}(\mathbf{X}_{K-1:K})$ has already a mixture representation of 7 components. In this example, we compute the approximate $\pi_{1:K|K}(\mathbf{X}_{1:K})$ by pruning mixture components with negligible weights. Fig. 4b, 4c and 4d show the estimated sets of trajectories in descending order of posterior probability. The results show that the estimate with unbroken trajectories has the highest weight. With the considered multi-object dynamic model, the set of trajectories with the highest posterior probability closely matches the true trajectories. The other sets of trajectories (with less probability) do not properly link one of the trajectories and, therefore, estimate an additional trajectory.

B. Simulation results

We present the results from a Monte Carlo simulation with 500 runs where the performance of the following multi-object filters/smothers are compared:

- 1) Track-oriented PMB filter [11], referred to as TO-PMB.
- 2) Variational PMB filter [43], referred to as V-PMB.
- 3) Multi-trajectory particle smoother with TO-PMB filtering densities, referred to as BS-TO-PMB.
- 4) Multi-trajectory particle smoother with V-PMB filtering densities, referred to as BS-V-PMB.
- 5) Trajectory PMBM filter for the set of all trajectories [18], referred to as T-PMBM.
- 6) Trajectory PMB filter for the set of all trajectories [21], referred to as T-PMB.
- 7) δ -GLMB filter with a multi-scan estimator [58] and RTS smoothing, referred to as GLMB.
- 8) Multi-scan GLMB with batch smoothing [23], referred to as M-GLMB.

We note that both 7) and 8) can be considered as implementations of the trajectory MBM₀₁ filter in [17].

We consider a two-dimensional scenario with 81 time steps where six initially well-separated objects move in close proximity to each other and thereafter separate, in the area $[-100 \text{ m}, 100 \text{ m}] \times [-100 \text{ m}, 100 \text{ m}]$. The true trajectories of the simulated scenario are illustrated in Fig. 5. We use a nearly constant velocity motion model with sampling period $T_s = 1$ s and single-object state transition density parameterised by

$$F = I_2 \otimes \begin{bmatrix} 1 & T_s \\ 0 & 1 \end{bmatrix}, \quad Q = \sigma_q^2 I_2 \otimes \begin{bmatrix} T_s^3/3 & T_s^2/2 \\ T_s^2/2 & T_s \end{bmatrix}$$

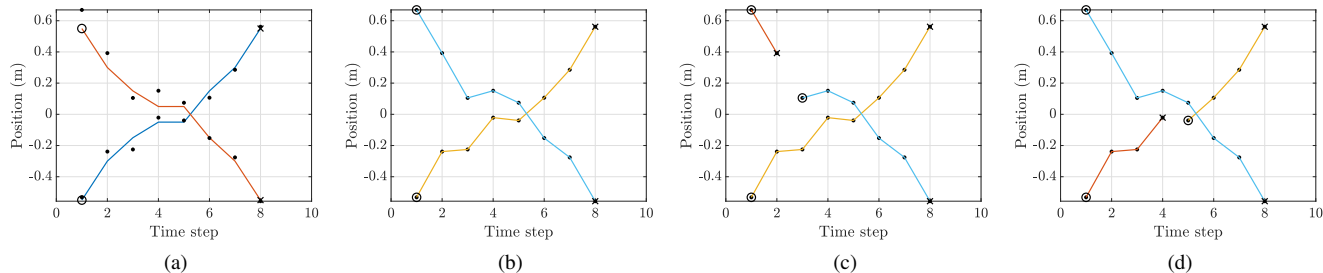


Fig. 4. Single-object Dirac deltas and true/estimated trajectories of a one-dimensional illustrative example. The single-object Dirac deltas are shown as black dots. Start/end positions of trajectories are marked by circles/crosses. Subfigure (a) shows the true trajectories. Subfigure (b), (c) and (d) show the estimated sets of trajectories in descending order of posterior probability; their posterior probabilities are 0.78, 0.15 and 0.03, respectively.

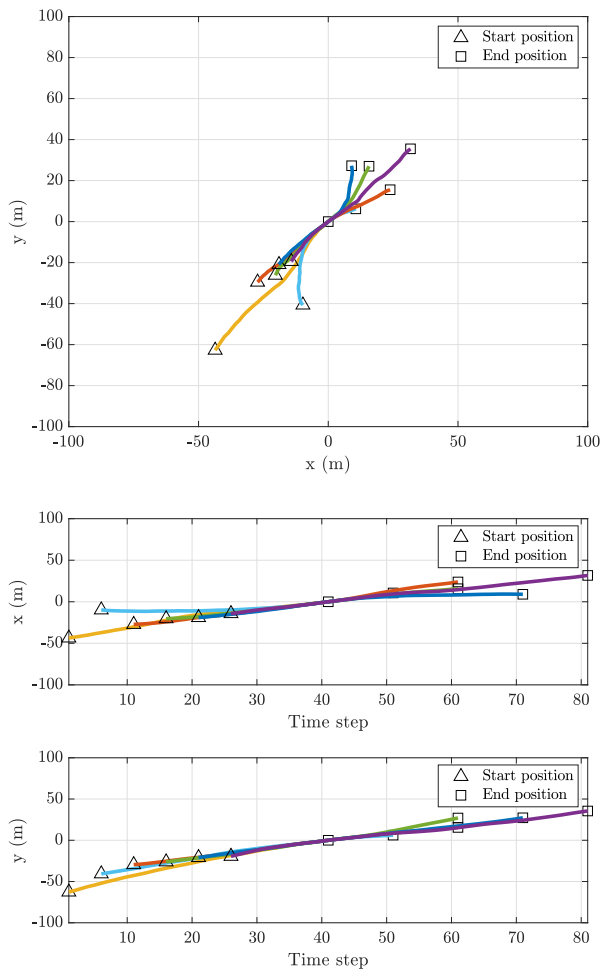


Fig. 5. True trajectories of the scenario with a high risk of track coalescence for one of 500 Monte Carlo runs. The figure on the top shows the two-dimensional trajectories, and how their decompositions into x and y dimensions vary over time is illustrated in the two figures on the bottom. Note that for different Monte Carlo runs, only the birth and death time of objects are deterministic. Start/end positions of trajectories are marked by triangle/square, respectively. There are six objects. They are born at time step 1, 6, 11, 16, 21, 26 and die at time step 41, 51, 61, 71, 81, respectively. At time step 41, one object dies when all objects are in close proximity.

where I_2 is an identity matrix, \otimes denotes the Kronecker product, and $\sigma_q = 0.1$. Each object survives with probability

$p^S = 0.98$. We also consider the point object measurement model where each object generates at most one measurement. The probability of detection is $p^D = 0.7$ and the clutter is uniformly distributed in the tracking area with Poisson rate $\lambda^C = 30$. The measurement model is also linear and Gaussian with observation matrix $I_2 \otimes [1 \ 0]$ and measurement noise covariance $\sigma_r^2 I_2$ where $\sigma_r = 1$.

For implementations with Poisson birth model, the Poisson birth intensity is a single Gaussian (c.f. (29)) with parameter $N_k^b = 1$, $w_k^{b,1} = 0.05$, $m_k^{b,1} = [-25, 1, -25, 1]^T$ and $P_k^{b,1} = \text{diag}(225, 1, 225, 1)$. The MB birth model used in GLMB and MS-GLMB contains a single Bernoulli with the same PHD as the Poisson birth process.

All the implementations use ellipsoidal gating (with gating size computed using the inverse-chi-squared distribution at probability 0.9999) to remove unlikely local hypotheses and Murty's algorithm to find the M -best global hypotheses with highest weight, with the only exception being MS-GLMB where the multi-scan data association problem is solved using Gibbs sampling. The maximum number of hypotheses is 100 for TO-PMB, V-PMB and T-PMB, and 1000 for T-PMBM and GLMB. For filters with Poisson birth, Bernoulli components with probability of existence smaller than 10^{-4} and Poisson components with weights smaller than 10^{-4} are pruned. For T-PMBM and T-PMB, Bernoulli components with probability of being alive at the current time step smaller than 10^{-4} are considered dead, and both filters are implemented without L -scan approximation, i.e., single-object states at different time steps are not considered independent. For T-PMBM and GLMB, we prune global hypotheses with weight smaller than 10^{-4} . For MS-GLMB, the smoothed multi-trajectory estimate obtained from GLMB is used to initialise the Markov chain and 1000 iterations are used in the multi-scan Gibbs sampler. For BS-TO-PMB and BS-V-PMB, the number of particles is 1000 and only a maximum of 100 global hypotheses with weight larger than 10^{-4} can be sampled.

For filters with Poisson birth, estimates are extracted from Bernoulli components with probability of existence $r \geq 0.5$. For T-PMBM, this is done for the global hypothesis with the highest weight. Also note that we only extract T-PMBM and T-PMB estimates at the last time step. For GLMB and M-GLMB, we also report estimates from the global hypothesis with the highest weight. For BS-TO-PMB and BS-V-PMB,

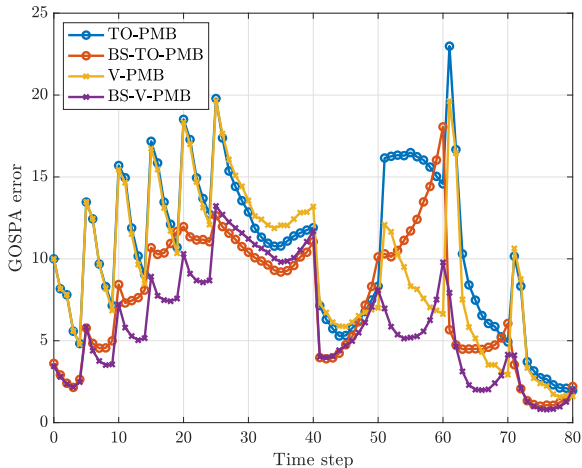


Fig. 6. GOSPA error versus time for the scenario with a high risk of track coalescence. The two multi-trajectory particle smoothers outperform their corresponding multi-object filters in terms of GOSPA error at almost all the time steps.

TABLE I

AVERAGE GOSPA ERROR AND ITS DECOMPOSITION WHERE LOCALISATION REFERS TO THE LOCALISATION ERROR NORMALISED BY THE ESTIMATED CARDINALITY

	GOSPA	Localisation	Missed	False
TO-PMB	859.5	93.6	358.5	196.0
BS-TO-PMB	606.8	64.3	272.3	119.2
V-PMB	772.0	83.3	363.8	140.0
BS-V-PMB	492.3	55.7	226.7	78.6

we report the particle (set of trajectories estimate) with the highest likelihood accumulated over time, see Appendix G for implementation details.

The multi-object state estimation performance is evaluated using the generalised optimal sub-pattern assignment (GOSPA) metric [63] with $\alpha = 2$, $c = 20$ and $p = 1$. Given a metric $d_b(\cdot, \cdot)$ in \mathbb{R}^{n_x} , a scalar $c > 0$, and a scalar p with $1 \leq p < \infty$, the GOSPA metric ($\alpha = 2$) between sets \mathbf{x} and \mathbf{y} is [63, Proposition 1]

$$d(\mathbf{x}, \mathbf{y}) = \left[\min_{\theta \in \Gamma} \left(\sum_{(i,j) \in \theta} d_b^p(x_i, y_j) + \frac{c^p}{2} (|\mathbf{x}| + |\mathbf{y}| - 2|\theta|) \right) \right]^{\frac{1}{p}}$$

where θ is an assignment set between sets $\{1, \dots, |\mathbf{x}|\}$ and $\{1, \dots, |\mathbf{y}|\}$, and Γ is the set of all possible assignment sets. In addition, the multi-trajectory estimation performance is evaluated using the linear programming (LP) metric for sets of trajectories [64] with $c = 20$, $p = 1$ and $\gamma = 2$, which is an extension of GOSPA to sets of trajectories.

We first analyse how the two multi-trajectory smoothers can improve multi-object estimation performance with respect to the forward filters. For TO-PMB, BS-TO-PMB, V-PMB and BS-V-PMB, the GOSPA error versus time is shown in Fig. 6, and the average GOSPA error and its decomposition into localisation error, missed detection error and false detection error are presented in Table I. It can be seen from Fig. 6

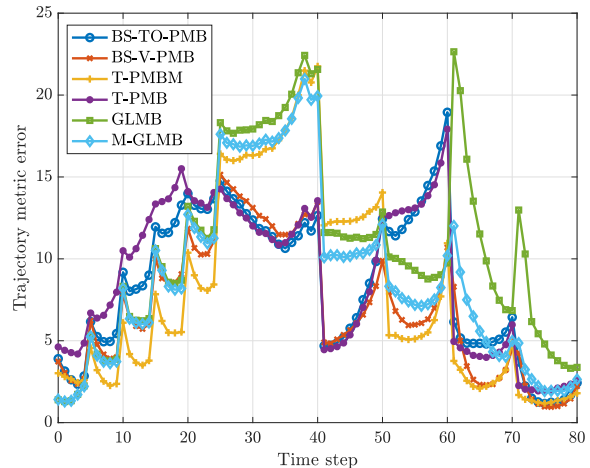


Fig. 7. LP trajectory metric error versus time for the scenario with a high risk of track coalescence. BS-V-PMB has the best overall performance.

TABLE II

AVERAGE LP TRAJECTORY METRIC ERROR AND ITS DECOMPOSITION

	Total	Localisation	Missed	False	Switch
BS-TO-PMB	688.2	277.9	274.9	118.3	17.2
BS-V-PMB	564.4	242.6	225.2	79.0	17.7
T-PMBM	640.5	243.9	356.2	22.9	17.5
T-PMB	722.3	280.3	237.8	185.9	18.3
GLMB	867.0	285.9	329.8	232.1	19.2
M-GLMB	724.3	262.5	324.5	116.5	18.8

that TO-PMB shows the largest estimation error when objects moving in close proximity begin to separate, a problem known as track coalescence commonly observed in JPDAF. As a comparison, V-PMB resolves the coalescence by using a more accurate MB approximation method [43]. Both BS-TO-PMB and BS-V-PMB outperform their corresponding forward filters in terms of localisation error, missed and false detections by a large margin. Between these two smoothers, BS-V-PMB has better estimation performance than BS-TO-PMB.

We proceed to analyse the trajectory estimation performance of different implementations. The LP trajectory metric error versus time for all the implementations that estimate trajectories is shown in Fig. 7, and the numerical values of the average LP trajectory metric error are presented in Table II. On the whole, BS-V-PMB has the best estimation performance averaged over different time steps, followed by T-PMBM. In particular, BS-V-PMB has the best performance when objects are in close proximity, whereas T-PMBM has the best performance on initiating and terminating trajectories. In principle, T-PMBM will produce optimal trajectory estimates if it is implemented without approximation, and we apply an optimal estimator (e.g. in the sense of minimising the mean trajectory metric error). However, in practice, we use pruning and a suboptimal estimator. When objects are well-spaced, only hypotheses with negligible weight are pruned, so the performance of T-PMBM is effectively optimal. With closely-spaced objects, there are many feasible hypotheses which cannot be effectively enumerated, and the additional ability of BS-V-PMB to reason over the entire sequence is

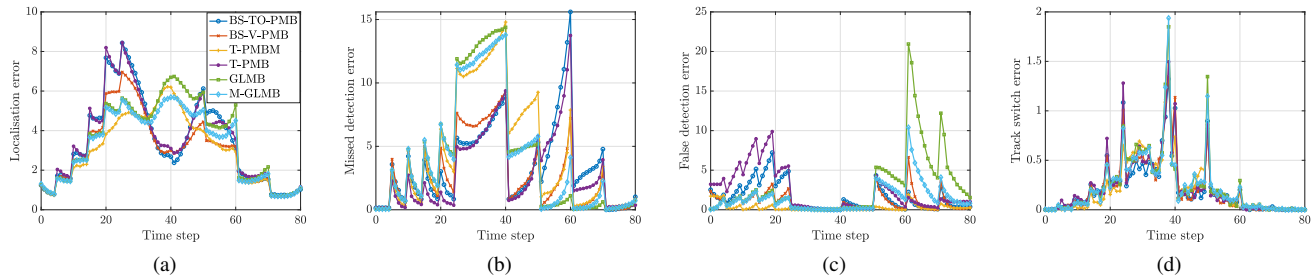


Fig. 8. Decompositions of trajectory metric error versus time for the scenario with a high risk of track coalescence.

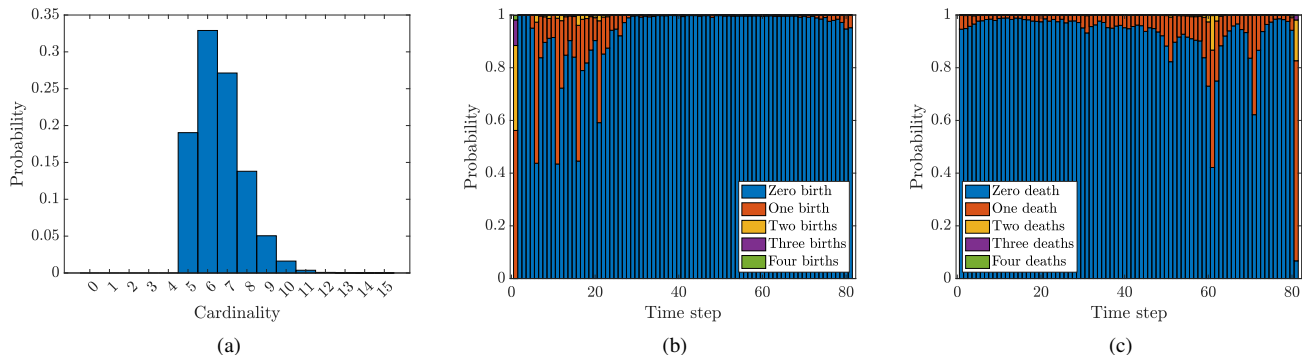


Fig. 9. Some statistics (averaged over 500 Monte Carlo runs) of the particle representation of the multi-trajectory density $\pi_{1:K|K}(\mathbf{X})$ obtained from BS-V-PMB. The computation of these statistics is given in Appendix F. Subfigure (a) shows the cardinality distribution of the set \mathbf{X} of trajectories in the time interval $1 : K$. Subfigure (b) and (c), respectively, show the cardinality distributions of trajectory birth and death at different time steps where the probability of a specific birth/death event is represented by the length of its corresponding bar.

clearly evident. BS-TO-PMB outperforms T-PMB, with the latter being an efficient approximation of T-PMBM. GLMB is less accurate than the other implementations using Poisson birth model due to their less efficient representation of the multi-object posterior [65], even though for the considered scenario where at most one object is born at a time it is more beneficial to use a Bernoulli birth model. M-GLMB has better performance than GLMB by improving the GLMB estimates using a multi-scan Gibbs sampler with batch smoothing.

The decompositions of the LP trajectory metric error into localisation error, missed detection error, false detection error and track switch error are shown in Fig. 8. T-PMBM, GLMB and M-GLMB show large missed detection errors when objects are in close proximity. In this case, many global hypotheses in T-PMBM and GLMB can have non-negligible weights due to the high data association uncertainty, and capping the number of global hypotheses may result in larger approximation errors than merging multiple global hypotheses into one. This explains why T-PMBM has worse performance than its approximation T-PMB when objects are in close proximity (before separation). Moreover, T-PMBM has very small false detection error, whereas GLMB has difficulty in terminating trajectories of dead objects. T-PMB has less missed detection error but more false detection error than BS-TO-PMB. The track switch error of all the implementations becomes large when objects are in close proximity, and implementations with Poisson birth model have lower track switch error than the two GLMB implementations.

Fig. 9 shows the cardinality distributions of the estimated set of trajectories as well as the trajectory birth and death at different time steps, computed using the particle representations of the multi-trajectory density $\pi_{1:K|K}(\mathbf{X})$ in BS-V-PMB. As can be seen, these statistics, in general, well reflect the ground truth except at time step 41 when one object died. Specifically, Fig. 9b and Fig. 9c show that it is very likely that no object dies at time step 41.

The average execution times in seconds of a single run² (81 time steps) for $p^D = 0.7$, $\lambda^C = 30$, $\sigma_q = 0.1$, and $\sigma_r = 1$ are: 281.0 (BS-TO-PMB), 209.4 (BS-V-PMB), 171.0 (T-PMBM), 6.5 (T-PMB), 11.1 (GLMB), 8683.7 (M-GLMB). The fastest implementation is T-PMB, followed by GLMB. T-PMBM is slower than GLMB as we do not use the L-scan approximation [21]. BS-V-PMB is faster than BS-TO-PMB even though V-PMB is slower than TO-PMB. This is due to the fact that the computational bottleneck of BS-V-PMB and BS-TO-PMB is backward simulation and that it is faster to run backward simulation on the PMB filtering densities obtained using V-PMB for the considered scenario. M-GLMB is significantly slower than the other implementations as it solves an 81-scan data association problem.

We proceed to analyse the performance of the filters and smoothers with Poisson birth for different scene parameters, and the results are shown in Table III. In general, BS-V-PMB has the best trajectory estimation performance, followed by

²MATLAB implementation on 3.0 GHz Intel Core i5.

TABLE III
AVERAGE LP TRAJECTORY METRIC ERROR FOR DIFFERENT SCENE
PARAMETERS OF THE SCENARIO WITH A HIGH RISK OF TRACK
COALESCENCE

	BS-TO-PMB	BS-V-PMB	T-PMBM	T-PMB
No change	688.2	564.4	640.5	722.3
$\sigma_q = 0.5$	1096.4	915.9	926.1	1088.2
$\sigma_r = 0.5$	460.5	374.7	409.8	476.4
$p^D = 0.8$	559.0	457.2	505.9	582.6
$\lambda^C = 10$	642.2	511.7	565.9	668.3
$\lambda^C = 50$	712.1	599.5	693.2	752.2

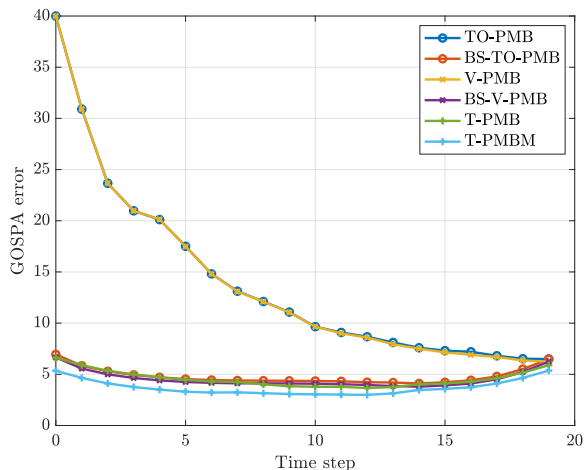


Fig. 10. GOSPA error versus time for the scenario with simultaneous object births. The two trajectory filters and the two multi-trajectory particle smoothers significantly outperform the two PMB filters based on sets of objects.

TABLE IV
AVERAGE GOSPA ERROR AND ITS DECOMPOSITION WHERE
LOCALISATION REFERS TO THE LOCALISATION ERROR NORMALISED BY
THE ESTIMATED CARDINALITY

	GOSPA	Localisation	Missed	False
TO-PMB	281.7	24.1	192.1	11.6
BS-TO-PMB	96.5	15.4	33.8	3.6
V-PMB	280.2	24.1	191.2	11.2
BS-V-PMB	90.7	15.4	29.8	1.8
T-PMBM	74.4	14.4	17.0	1.2
T-PMB	91.5	15.7	22.8	7.7

T-PMBM. As expected, if the scenario has lower signal-to-noise ratio, e.g., when the motion noise or clutter intensity increases, performance of all the filters/smoothers decreases. If the scenario has higher signal-to-noise ratio, e.g., measurement noise or clutter intensity decreases, or probability of detection increases, performance of all the filters/smoothers increases.

To further demonstrate the ability of the proposed multi-trajectory smoothers to infer object birth locations before first detection, we consider another scenario with 20 time steps where four objects are born at time step 1 and no object dies. Compared to the first scenario, here the probability of detection is $p^D = 0.5$ and the Poisson clutter rate is $\lambda^C = 5$. We compare the performance of different implementations with Poisson birth model. The Poisson birth intensity is parameterised by $N_k^b = 1$, $w_k^{b,1} = 0.02$, $x_k^{b,1} = [0, 0, 0, 0]^T$,

$P_k^{b,1} = \text{diag}(100, 2, 100, 2)^2$, and the initial Poisson intensity for undetected objects is set to $\lambda_{0|0}^u(x) = 4\mathcal{N}(x; x_k^{b,1}, P_k^{b,1})$.

The GOSPA error versus time for the considered scenario is shown in Fig. 10 and its decomposition is presented in Table IV. The results show that T-PMBM has the best estimation performance, and that TO-PMB and V-PMB have the worst estimation performance. The estimation performance of T-PMB, BS-TO-PMB and BS-V-PMB is similar, and it is slightly worse than T-PMBM. Due to the low detection probability, objects were usually detected a few time steps after they were born. This explains the high missed detection error of TO-PMB and V-PMB. The estimation of object states before first detection can be obtained by considering the posterior density on sets of trajectories, which captures all the information about the trajectories, including those of undetected objects. For T-PMBM and T-PMB, this information is explicitly carried over time via forward filtering, whereas for BS-TO-PMB and BS-V-PMB, this information is inferred from filtering densities of detected objects at later time steps and gradually recovered via backward smoothing.

At last, we note that the forward-backward smoothers BS-TO-PMB, BS-V-PMB and the trajectory filters T-PMBM, T-PMB are suitable for different applications. BS-TO-PMB and BS-V-PMB are offline methods and their backward smoothing steps do not utilise any information of the measurements and the multi-object measurement model, and therefore they are more suitable for offline trajectory analytics. T-PMBM and T-PMB are online methods, and they can also be applied to batch problems. The difference is that T-PMBM and T-PMB solve the data associations while filtering, whereas with backward simulation, we are not constrained to the previously solved data associations in the sense that future measurements can be utilised to improve trajectory estimation.

VII. CONCLUSIONS

In this paper, we have derived a multi-trajectory backward smoothing equation based on sets of trajectories. This allows us to leverage filters that do not keep trajectory information to compute the posterior density of sets of trajectories, and has important applications to offline trajectory analytics. In addition, we have proposed a multi-trajectory particle smoother using backward simulation for PMB filtering densities along with its tractable implementation based on ranked assignment. The simulation results show that the proposed methods have superior trajectory estimation performance compared to several state-of-the-art algorithms.

A follow-up work direction is developing an implementation that works for forward densities with particle representation. In addition, it would be interesting to study how to extract better estimates from the particle representation of multi-trajectory densities (28), e.g., by merging different particles in the multi-object trajectory space.

REFERENCES

- [1] Y. Bar-Shalom, X. R. Li, and T. Kirubarajan, *Estimation with applications to tracking and navigation: theory algorithms and software*. John Wiley & Sons, 2004.

- [2] S. Challa, M. R. Morelande, D. Mušicki, and R. J. Evans, *Fundamentals of object tracking*. Cambridge University Press, 2011.
- [3] F. Meyer, T. Kropfreiter, J. L. Williams, R. Lau, F. Hlawatsch, P. Braca, and M. Z. Win, "Message passing algorithms for scalable multitarget tracking," *Proceedings of the IEEE*, vol. 106, no. 2, pp. 221–259, 2018.
- [4] R. L. Streit, R. B. Angle, and M. Efe, *Analytic combinatorics in multiple object tracking*. Springer, 2021.
- [5] Y. Bar-Shalom, F. Daum, and J. Huang, "The probabilistic data association filter," *IEEE Control Systems Magazine*, vol. 29, no. 6, pp. 82–100, 2009.
- [6] S. S. Blackman, "Multiple hypothesis tracking for multiple target tracking," *IEEE Aerospace and Electronic Systems Magazine*, vol. 19, no. 1, pp. 5–18, 2004.
- [7] C. Chong, S. Mori, and D. B. Reid, "Forty years of multiple hypothesis tracking," *Journal of Advances in Information Fusion*, vol. 14, no. 2, pp. 131–151, 2019.
- [8] R. P. Mahler, *Statistical Multisource-Multitarget Information Fusion*. Artech House Norwood, MA, 2007.
- [9] —, "Multitarget Bayes filtering via first-order multitarget moments," *IEEE Transactions on Aerospace and Electronic systems*, vol. 39, no. 4, pp. 1152–1178, 2003.
- [10] —, "PHD filters of higher order in target number," *IEEE Transactions on Aerospace and Electronic systems*, vol. 43, no. 4, pp. 1523–1543, 2007.
- [11] J. L. Williams, "Marginal multi-Bernoulli filters: RFS derivation of MHT, JPDA, and association-based member," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 51, no. 3, pp. 1664–1687, 2015.
- [12] B.-T. Vo and B.-N. Vo, "Labeled random finite sets and multi-object conjugate priors," *IEEE Transactions on Signal Processing*, vol. 61, no. 13, pp. 3460–3475, 2013.
- [13] Á. F. García-Fernández, J. Grajal, and M. R. Morelande, "Two-layer particle filter for multiple target detection and tracking," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 49, no. 3, pp. 1569–1588, 2013.
- [14] Á. F. García-Fernández, M. R. Morelande, and J. Grajal, "Bayesian sequential track formation," *IEEE Transactions on Signal Processing*, vol. 62, no. 24, pp. 6366–6379, 2014.
- [15] R. Streit, "Analytic combinatorics and labeling in high level fusion and multihypothesis tracking," in *21st International Conference on Information Fusion (FUSION)*. IEEE, 2018, pp. 1–5.
- [16] E. H. Aoki, P. K. Mandal, L. Svensson, Y. Boers, and A. Bagchi, "Labeling uncertainty in multitarget tracking," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 52, no. 3, pp. 1006–1020, 2016.
- [17] Á. F. García-Fernández, L. Svensson, and M. R. Morelande, "Multiple target tracking based on sets of trajectories," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 56, no. 3, pp. 1685–1707, 2019.
- [18] K. Granström, L. Svensson, Y. Xia, J. Williams, and Á. F. García-Fernández, "Poisson multi-Bernoulli mixture trackers: Continuity through random finite sets of trajectories," in *21st International Conference on Information Fusion (FUSION)*. IEEE, 2018, pp. 1–8.
- [19] Á. F. García-Fernández and L. Svensson, "Trajectory PHD and CPHD filters," *IEEE Transactions on Signal Processing*, vol. 67, no. 22, pp. 5702–5714, 2019.
- [20] Y. Xia, K. Granström, L. Svensson, Á. F. García-Fernández, and J. L. Williams, "Multi-scan implementation of the trajectory Poisson multi-Bernoulli mixture filter," *Journal of Advances in Information Fusion*, vol. 14, no. 2, pp. 213–235, 2019.
- [21] Á. F. García-Fernández, L. Svensson, J. L. Williams, Y. Xia, and K. Granström, "Trajectory Poisson multi-Bernoulli filters," *IEEE Transactions on Signal Processing*, vol. 68, pp. 4933–4945, 2020.
- [22] —, "Trajectory multi-Bernoulli filters for multi-target tracking based on sets of trajectories," in *IEEE 23rd International Conference on Information Fusion (FUSION)*. IEEE, 2020, pp. 1–8.
- [23] B.-N. Vo and B.-T. Vo, "A multi-scan labeled random finite set model for multi-object state estimation," *IEEE Transactions on Signal Processing*, vol. 67, no. 19, pp. 4948–4963, 2019.
- [24] S. Särkkä, *Bayesian filtering and smoothing*. Cambridge University Press, 2013, no. 3.
- [25] G. Kitagawa, "The two-filter formula for smoothing and an implementation of the Gaussian-sum smoother," *Annals of the Institute of Statistical Mathematics*, vol. 46, no. 4, pp. 605–623, 1994.
- [26] D. J. Lee and M. E. Campbell, "Smoothing algorithm for nonlinear systems using Gaussian mixture models," *Journal of Guidance, Control, and Dynamics*, vol. 38, no. 8, pp. 1438–1451, 2015.
- [27] A. S. Rahmathullah, L. Svensson, and D. Svensson, "Merging-based forward-backward smoothing on Gaussian mixtures," in *17th International Conference on Information Fusion (FUSION)*. IEEE, 2014, pp. 1–8.
- [28] —, "Two-filter Gaussian mixture smoothing with posterior pruning," in *17th International Conference on Information Fusion (FUSION)*. IEEE, 2014, pp. 1–8.
- [29] M. P. Balenzuela, J. Dahlin, N. Bartlett, A. G. Wills, C. Renton, and B. Ninness, "Accurate Gaussian mixture model smoothing using a two-filter approach," in *IEEE Conference on Decision and Control (CDC)*. IEEE, 2018, pp. 694–699.
- [30] R. Chakravorty and S. Challa, "Augmented state integrated probabilistic data association smoothing for automatic track initiation in clutter," *J. Adv. Inf. Fusion*, vol. 1, no. 1, pp. 63–74, 2006.
- [31] T. L. Song and D. Mušicki, "Smoothing innovations and data association with IPDA," *Automatica*, vol. 48, no. 7, pp. 1324–1329, 2012.
- [32] D. Mušicki, T. L. Song, and T. H. Kim, "Smoothing multi-scan target tracking in clutter," *IEEE Transactions on Signal Processing*, vol. 61, no. 19, pp. 4740–4752, 2013.
- [33] D. Clark, "Joint target-detection and tracking smoothers," in *Signal Processing, Sensor Fusion, and Target Recognition XVIII*, vol. 7336. International Society for Optics and Photonics, 2009, p. 73360G.
- [34] B.-T. Vo, D. Clark, B.-N. Vo, and B. Ristic, "Bernoulli forward-backward smoothing for joint target detection and tracking," *IEEE Transactions on Signal Processing*, vol. 59, no. 9, pp. 4473–4477, 2011.
- [35] B.-N. Vo, B.-T. Vo, and R. P. Mahler, "Closed-form solutions to forward-backward smoothing," *IEEE Transactions on Signal Processing*, vol. 60, no. 1, pp. 2–17, 2011.
- [36] H. E. Rauch, F. Tung, and C. T. Striebel, "Maximum likelihood estimates of linear dynamic systems," *AIAA journal*, vol. 3, no. 8, pp. 1445–1450, 1965.
- [37] A. Mahalanabis, B. Zhou, and N. Bose, "Improved multi-target tracking in clutter by PDA smoothing," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 26, no. 1, pp. 113–121, 1990.
- [38] W. Koch, "Fixed-interval retrodiction approach to Bayesian IMM-MHT for maneuvering multiple targets," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 36, no. 1, pp. 2–14, 2000.
- [39] B. Chen and J. K. Tugnait, "Tracking of multiple maneuvering targets in clutter using IMM/JPDA filtering and fixed-lag smoothing," *Automatica*, vol. 37, no. 2, pp. 239–249, 2001.
- [40] A. S. Rahmathullah, R. Selvan, and L. Svensson, "A batch algorithm for estimating trajectories of point targets using expectation maximization," *IEEE Transactions on Signal Processing*, vol. 64, no. 18, pp. 4792–4804, 2016.
- [41] M. Briers, A. Doucet, and S. Maskell, "Smoothing algorithms for state-space models," *Annals of the Institute of Statistical Mathematics*, vol. 62, no. 1, p. 61, 2010.
- [42] L. Svensson, D. Svensson, M. Guerriero, and P. Willett, "Set JPDA filter for multitarget tracking," *IEEE Transactions on Signal Processing*, vol. 59, no. 10, pp. 4677–4691, 2011.
- [43] J. L. Williams, "An efficient, variational approximation of the best fitting multi-Bernoulli filter," *IEEE Transactions on Signal Processing*, vol. 63, no. 1, pp. 258–273, 2015.
- [44] D. Svensson, L. Svensson, M. Guerriero, D. F. Crouse, and P. Willett, "The multitarget set JPDA filter with target identity," in *Signal Processing, Sensor Fusion, and Target Recognition XX*, vol. 8050. International Society for Optics and Photonics, 2011, p. 805010.
- [45] N. Nandakumar, K. Punithakumar, and T. Kirubarajan, "Improved multitarget tracking using probability hypothesis density smoothing," in *Signal and Data Processing of Small Targets*, p. 6699. International Society for Optics and Photonics, 2007, p. 66990M.
- [46] D. E. Clark, "First-moment multi-object forward-backward smoothing," in *13th International Conference on Information Fusion*. IEEE, 2010, pp. 1–6.
- [47] N. Nadarajah, T. Kirubarajan, T. Lang, M. McDonald, and K. Punithakumar, "Multitarget tracking using probability hypothesis density smoothing," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 47, no. 4, pp. 2344–2360, 2011.
- [48] R. P. Mahler, B.-T. Vo, and B.-N. Vo, "Forward-backward probability hypothesis density smoothing," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 48, no. 1, pp. 707–728, 2012.
- [49] P. Feng, W. Wang, S. M. Naqvi, and J. Chambers, "Adaptive retrodiction particle PHD filter for multiple human tracking," *IEEE Signal Processing Letters*, vol. 23, no. 11, pp. 1592–1596, 2016.
- [50] D. Li, C. Hou, and D. Yi, "Multi-Bernoulli smoother for multi-target tracking," *Aerospace Science and Technology*, vol. 48, pp. 234–245, 2016.

- [51] S. Nagappa, E. D. Delande, D. E. Clark, and J. Houssineau, "A tractable forward-backward CPHD smoother," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 53, no. 1, pp. 201–217, 2017.
- [52] R. Streit, "Interval/smoothing filters for multiple object tracking via analytic combinatorics," in *20th International Conference on Information Fusion (Fusion)*. IEEE, 2017, pp. 1–8.
- [53] M. Beard, B. T. Vo, and B.-N. Vo, "Generalised labelled multi-Bernoulli forward-backward smoothing," in *19th International Conference on Information Fusion (FUSION)*. IEEE, 2016, pp. 688–694.
- [54] R. Liu, H. Fan, T. Li, and H. Xiao, "A computationally efficient labeled multi-Bernoulli smoother for multi-target tracking," *Sensors*, vol. 19, no. 19, p. 4226, 2019.
- [55] G. Liang, Q. Li, B. Qi, and L. Qiu, "Multitarget tracking using one time step lagged delta-generalized labeled multi-Bernoulli smoothing," *IEEE Access*, vol. 8, pp. 28 242–28 256, 2020.
- [56] Y. Xia, L. Svensson, Á. F. García-Fernández, K. Granström, and J. L. Williams, "Backward simulation for sets of trajectories," in *IEEE 23rd International Conference on Information Fusion (FUSION)*. IEEE, 2020, pp. 1–8.
- [57] F. Lindsten and T. B. Schön, "Backward simulation methods for Monte Carlo statistical inference," *Foundations and Trends® in Machine Learning*, vol. 6, no. 1, pp. 1–143, 2013.
- [58] T. T. D. Nguyen and D. Y. Kim, "GLMB tracker with partial smoothing," *Sensors*, vol. 19, no. 20, p. 4419, 2019.
- [59] K. Granström, L. Svensson, Y. Xia, J. Williams, and Á. F. García-Fernández, "Poisson multi-Bernoulli mixtures for sets of trajectories," *arXiv preprint arXiv:1912.08718*, 2019.
- [60] H. A. Blom and Y. Bar-Shalom, "The interacting multiple model algorithm for systems with Markovian switching coefficients," *IEEE Transactions on Automatic Control*, vol. 33, no. 8, pp. 780–783, 1988.
- [61] Á. F. García-Fernández, J. L. Williams, L. Svensson, and Y. Xia, "A Poisson multi-Bernoulli mixture filter for coexisting point and extended targets," *IEEE Transactions on Signal Processing*, vol. 69, pp. 2600–2610, 2021.
- [62] D. F. Crouse, "On implementing 2D rectangular assignment algorithms," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 52, no. 4, pp. 1679–1696, 2016.
- [63] A. S. Rahmathullah, Á. F. García-Fernández, and L. Svensson, "Generalized optimal sub-pattern assignment metric," in *Proceedings of International Conference on Information Fusion*. IEEE, 2017, pp. 1–8.
- [64] Á. F. García-Fernández, A. S. Rahmathullah, and L. Svensson, "A metric on the space of finite sets of trajectories for evaluation of multi-target tracking algorithms," *IEEE Transactions on Signal Processing*, vol. 68, pp. 3917–3928, 2020.
- [65] Á. F. García-Fernández, J. L. Williams, K. Granström, and L. Svensson, "Poisson multi-Bernoulli mixture filter: direct derivation and implementation," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 54, no. 4, pp. 1883–1901, 2018.

Supplementary Materials

APPENDIX A PROOF OF THEOREM 1

We prove Theorem 1 by induction³. The base case one-step prediction with $\gamma = \eta + 1$ has been proved in [17]. We proceed to show that if (11) and (12) hold, then the $(\gamma + 1 - \eta)$ -step predicted multi-trajectory density of $\pi_{\alpha:\eta|k}(\cdot)$ is also of the form (11) and (12). We define $\mathbf{W}^{\gamma+1}$ as the set of trajectories born at time step $\gamma + 1$. Then the one-step predicted multi-trajectory density of $\pi_{\alpha:\gamma|k}(\mathbf{X}_{\alpha:\gamma})$ is

$$\begin{aligned} \pi_{\alpha:\gamma+1|k}(\mathbf{X}_{\alpha:\gamma+1}) &= \pi_{\alpha:\gamma|k}(\mathbf{X}_{\alpha:\gamma})\pi_{\gamma+1:\gamma+1}(\mathbf{W}^{\gamma+1}) \\ &\times \prod_{(t,x^{1:\nu}) \in \mathbf{X}_{\alpha:\gamma+1}^{\gamma}} \left[(1 + p^S(x^\nu) (\delta_{\gamma-t+2}[\nu] - 1)) \right. \\ &\quad \left. \times \prod_{\ell=\gamma-t+1}^{\nu-1} g(x^{\ell+1}|x^\ell) p^S(x^\ell) \right]. \end{aligned} \quad (37)$$

We further write $\mathbf{X}_{\alpha:\gamma+1}^{\gamma} = \mathbf{Y} \uplus \mathbf{V}$ where $\mathbf{Y} \subseteq \mathbf{X}_{\alpha:\gamma+1}^{\eta}$ is the set of trajectories present at time step η and $\mathbf{V} \subseteq \mathbf{W}$ is the set of trajectories that appeared after time step η (by assumption $\eta < \gamma$). This allows us to separate the product over $\mathbf{X}_{\alpha:\gamma+1}^{\gamma}$ in (37) into two products over \mathbf{Y} and \mathbf{V} , respectively. By plugging (11) and (12) into (37) and combining these two products with the product over $\mathbf{X}_{\alpha:\gamma}^{\eta}$ in (11) and the product over \mathbf{W} in (12), respectively, we obtain

$$\begin{aligned} \pi_{\alpha:\gamma+1|k}(\mathbf{X}_{\alpha:\gamma+1}) &= \pi_{\alpha:\eta|k}(\mathbf{X}_{\alpha:\eta})\pi_{\eta+1:\gamma+1}(\mathbf{W} \uplus \mathbf{W}^{\gamma+1}) \\ &\times \prod_{(t,x^{1:\nu}) \in \mathbf{X}_{\alpha:\gamma+1}^{\eta}} \left[(1 + p^S(x^\nu) (\delta_{\gamma-t+2}[\nu] - 1)) \right. \\ &\quad \left. \times \prod_{\ell=\eta-t+1}^{\nu-1} g(x^{\ell+1}|x^\ell) p^S(x^\ell) \right], \end{aligned} \quad (38)$$

$$\begin{aligned} \pi_{\eta+1:\gamma+1}(\mathbf{W} \uplus \mathbf{W}^{\gamma+1}) &= \prod_{\ell=\eta+1}^{\gamma+1} \beta(\tau^\ell(\mathbf{W}^\ell)) \\ &\times \prod_{(t,x^{1:\nu}) \in \mathbf{W}} \left[(1 + p^S(x^\nu) (\delta_{\gamma-t+2}[\nu] - 1)) \right. \\ &\quad \left. \times \prod_{\ell=1}^{\nu-1} g(x^{\ell+1}|x^\ell) p^S(x^\ell) \right]. \end{aligned} \quad (39)$$

We can then observe that (38) has the same form as (11), and that (39) has the same form as (12).

This finishes the proof of Theorem 1.

APPENDIX B PROOF OF COROLLARY 1.1

We observe that the only factor in (11) that depends on time step α is $\pi_{\alpha:\eta|k}(\mathbf{X}_{\alpha:\eta})$. This means that the quotient

$$\frac{\pi_{\alpha:\gamma|k}(\mathbf{X}_{\alpha:\gamma})}{\pi_{\alpha:\eta|k}(\mathbf{X}_{\alpha:\eta})}$$

³A direct proof using sets integrals can be found in [56, Appendix A].

does not depend on time step α , and therefore it holds that

$$\frac{\pi_{k:\gamma|k}(\mathbf{X}_{k:\gamma})}{\pi_{k:\eta|k}(\mathbf{X}_{k:\eta})} = \frac{\pi_{k+1:\gamma|k}(\mathbf{X}_{k+1:\gamma})}{\pi_{k+1:\eta|k}(\mathbf{X}_{k+1:\eta})}. \quad (40)$$

Setting $\eta = k + 1$ in (40) gives

$$\frac{\pi_{k:\gamma|k}(\mathbf{X}_{k:\gamma})}{\pi_{k:k+1|k}(\mathbf{X}_{k:k+1})} = \frac{\pi_{k+1:\gamma|k}(\mathbf{X}_{k+1:\gamma})}{f_{k+1|k}(\tau^{k+1}(\mathbf{X}_{k+1:k+1}))}, \quad (41)$$

which can be rearranged to obtain (13).

This finishes the proof of Corollary 1.1.

APPENDIX C PROOF OF THEOREM 2

Let \mathbf{X} and \mathbf{Y} be the set of trajectories in the time interval $k : K$ and $k + 1 : K$, respectively. Applying the total probability theorem and Bayes' rule gives

$$\begin{aligned} \pi_{k:K|K}(\mathbf{X}) &= \int \pi_{k:K|K}(\mathbf{X}, \mathbf{Y}) \delta \mathbf{Y}, \\ &= \int \pi_{k+1:K|K}(\mathbf{Y}) \pi_{k:K|K}(\mathbf{X}|\mathbf{Y}) \delta \mathbf{Y} \end{aligned} \quad (42)$$

where $\pi_{k:K|K}(\mathbf{X}|\mathbf{Y})$ defines the backward transition density from \mathbf{Y} to \mathbf{X} conditioned on the sequence of sets of measurements up to and including time step K . Assume that \mathbf{X} is independent of measurements $(\mathbf{z}_{k+1}, \dots, \mathbf{z}_K)$ that are in the future given \mathbf{Y} :

$$\pi_{k:K|K}(\mathbf{X}|\mathbf{Y}) = \pi_{k:K|k}(\mathbf{X}|\mathbf{Y}). \quad (43)$$

Then we have

$$\begin{aligned} \pi_{k:K|K}(\mathbf{X}) &= \int \pi_{k+1:K|K}(\mathbf{Y}) \pi_{k:K|K}(\mathbf{X}|\mathbf{Y}) \delta \mathbf{Y} \\ &= \pi_{k:K|k}(\mathbf{X}) \int \pi_{k+1:K|K}(\mathbf{Y}) \frac{\pi_{k:K|k}(\mathbf{X}|\mathbf{Y})}{\pi_{k:K|k}(\mathbf{X})} \delta \mathbf{Y}. \end{aligned} \quad (44)$$

Bayes' rule then yields

$$\pi_{k:K|K}(\mathbf{X}) = \pi_{k:K|k}(\mathbf{X}) \int \pi_{k+1:K|K}(\mathbf{Y}) \frac{\pi_{k+1:K|k}(\mathbf{Y}|\mathbf{X})}{\pi_{k+1:K|k}(\mathbf{Y})} \delta \mathbf{Y} \quad (45)$$

where $\pi_{k+1:K|k}(\mathbf{Y}|\mathbf{X})$ defines the transition density from \mathbf{X} to \mathbf{Y} , which is a multi-trajectory Dirac delta. The integral over \mathbf{Y} in (45) can then be cancelled out by applying the prediction equation for sets of trajectories [17, Eq. (8)], which gives us

$$\pi_{k:K|K}(\mathbf{X}) = \frac{\pi_{k:K|k}(\mathbf{X})\pi_{k+1:K|K}(\mathbf{X}_{k+1:K})}{\pi_{k+1:K|k}(\mathbf{X}_{k+1:K})}. \quad (46)$$

Applying Corollary 1.1 on (46) yields

$$\frac{\pi_{k:K|k}(\mathbf{X})}{\pi_{k+1:K|k}(\mathbf{X}_{k+1:K})} = \frac{\pi_{k:k+1|k}(\mathbf{X}_{k:k+1})}{f_{k+1|k}(\tau^{k+1}(\mathbf{X}_{k+1:k+1}))}. \quad (47)$$

The proof is finished by plugging (47) into (46).

APPENDIX D
PROOF OF LEMMA 3

We first rewrite $\pi_{k:K|K}(\mathbf{X}|\mathbf{Y})$ using (43) and Bayes' rule:

$$\pi_{k:K|K}(\mathbf{X}|\mathbf{Y}) = \frac{\pi_{k:K|k}(\mathbf{X})\pi_{k+1:K|k}(\mathbf{Y}|\mathbf{X})}{\pi_{k+1:K|k}(\mathbf{Y})} \quad (48)$$

where $\pi_{k+1:K|k}(\mathbf{Y}|\mathbf{X})$ defines the transition density from \mathbf{X} to \mathbf{Y} , which is a multi-trajectory Dirac delta. We then apply Corollary 1.1, which gives

$$\pi_{k:K|K}(\mathbf{X}|\mathbf{Y}) = \frac{\pi_{k:k+1|k}(\mathbf{X}_{k:k+1})\delta_{\mathbf{Y}}(\mathbf{X}_{k+1:K})}{f_{k+1|k}(\tau^{k+1}(\mathbf{Y}_{k+1:k+1}))}. \quad (49)$$

As $f_{k+1|k}(\tau^{k+1}(\mathbf{Y}_{k+1:k+1}))$ does not depend on \mathbf{X} , we can further express $\pi_{k:K|k}(\mathbf{X}|\mathbf{Y})$ as (15).

This finishes the proof of Lemma 3.

APPENDIX E
PROOF OF THEOREM 4

We first give the explicit expression of the multi-trajectory density $\pi_{k:k+1|k}(\mathbf{X}_{k:k+1})$. Given the PMB filtering density at time step k (cf. (10)) and the multi-trajectory dynamic model described in Section II-C with Poisson birth density (9), the predicted multi-trajectory density $\pi_{k:k+1|k}(\mathbf{X}_{k:k+1})$ is a PMB [21, Lemma 4]

$$\pi_{k:k+1|k}(\mathbf{X}_{k:k+1}) = \sum_{\substack{\mathfrak{J} \\ \mathfrak{J} \cup \mathbf{P} = \mathbf{X}_{k:k+1}}} \pi_{k:k+1|k}^{\mathfrak{P}}(\mathbf{P}) \prod_{i=1}^{n_{k:k+1|k}} \left[\pi_{k:k+1|k}^i(\mathbf{X}^i) \right] \quad (50)$$

with $n_{k:k+1|k} = n_{k|k}$, where $\pi_{k:k+1|k}^{\mathfrak{P}}(\cdot)$ is of the form (3) with intensity

$$\lambda_{k:k+1|k}^u(t, x^{1:\nu}) = \delta_{k+1}[t]\delta_1[\nu]\lambda_{k+1}^B(x^1) + \left\langle \lambda_{k|k}^u(y^1) \delta_k[t']\delta_1[\nu'], g^{k+1}(t, x^{1:\nu}|t', y^{1:\nu'}) \right\rangle, \quad (51)$$

and $\pi_{k:k+1|k}^i(\mathbf{X}^i)$ is of the form (4), parameterised by

$$r_{k:k+1|k}^i = r_{k|k}^i, \quad (52)$$

$$p_{k:k+1|k}^i(X) = \left\langle p_{k|k}^i(y^1) \delta_k[t']\delta_1[\nu'], g^{k+1}(X|t', y^{1:\nu'}) \right\rangle. \quad (53)$$

where the single trajectory transition density $g^{k+1}(\cdot|\cdot)$ is given by (6).

Next, we elaborate on why the multi-trajectory Dirac delta $\delta_{\mathbf{Y}}(\mathbf{X}_{k+1:K})$ can be seen as a standard multi-object measurement model [8] with characteristics specified in Section IV-B. We note that the multi-trajectory Dirac delta $\delta_{\mathbf{Y}}(\mathbf{X}_{k+1:K})$ can be written as a trajectory MB where each trajectory Bernoulli component is parameterised by probability of existence one and a Dirac delta single-trajectory density, i.e.,

$$\delta_{\mathbf{Y}}(\mathbf{X}_{k+1:K}) = \sum_{\substack{\mathfrak{J} \\ \mathfrak{J} \cup \mathbf{P} = \mathbf{X}_{k+1:K}}} \prod_{i=1}^{n_{k+1:K}} \delta_{\{Y^i\}}(\mathbf{X}^i) \quad (54)$$

where for trajectories that did not exist in the time interval $k+1 : K$, they are implicitly represented by trajectory Bernoulli

components with zero probability of existence. Therefore, the multi-trajectory Dirac delta $\delta_{\mathbf{Y}}(\mathbf{X}_{k+1:K})$ can be understood as a standard measurement model for sets of trajectories [21]:

- Each trajectory $X = (t, x^{1:\nu}) \in \mathbf{X}_{k:K}$ is detected with probability

$$p^D(X) = \begin{cases} 0, & t = k \text{ and } \nu = 1 \\ 1, & \text{otherwise} \end{cases} \quad (55)$$

and if detected, it generates a measurement Y with density $\delta_Y(X_{k+1:K})$.

- Poisson clutter intensity $\lambda^C(\cdot) = 0$.

Note that if a trajectory X in the time interval $k : K$ did not exist in the time interval $k+1 : K$, then it must have start time $t = k$ and length $\nu = 1$.

Having established the analogy between the backward kernel (15) for PMB filtering densities and the trajectory PMB update, the explicit expression of the trajectory PMBM backward kernel (19) can be straightforwardly obtained by plugging the trajectory PMB prior (50) and measurement model (54) into the trajectory PMB update equations [21, Lemma 5]. One thing to be noted is that the set of trajectories appeared after time step $k+1$ remains unaltered. This will be further elaborated.

We write $\mathbf{X} = \mathbf{X}' \uplus \mathbf{D}'$ as the disjoint union of the set \mathbf{X}' of trajectories present at time step k or $k+1$ and the set \mathbf{D}' of trajectories of objects that appeared after time step $k+1$. We also write $\mathbf{Y} = \mathbf{Y}' \uplus \mathbf{D}$ as the disjoint union of the set \mathbf{Y}' of trajectories of objects that existed at time step $k+1$ and the set \mathbf{D} of trajectories of objects that appeared after time step $k+1$. It holds that $\mathbf{X}_{k+1:K} = \mathbf{X}'_{k+1:K} \uplus \mathbf{D}'$ by construction, and that the density of $\mathbf{X}_{k+1:K}$ conditioned on \mathbf{Y} is zero unless $\mathbf{D} = \mathbf{D}'$. Then we have

$$\delta_{\mathbf{Y}' \uplus \mathbf{D}}(\mathbf{X}'_{k+1:K} \uplus \mathbf{D}') = \delta_{\mathbf{Y}'}(\mathbf{X}'_{k+1:K}) \delta_{\mathbf{D}}(\mathbf{D}'), \quad (56)$$

and the backward kernel (15) becomes

$$\pi_{k:K|K}(\mathbf{X}' \uplus \mathbf{D}' | \mathbf{Y}' \uplus \mathbf{D}) \propto \pi_{k:k+1|k}(\mathbf{X}'_{k:k+1}) \delta_{\mathbf{Y}'}(\mathbf{X}'_{k+1:K}) \delta_{\mathbf{D}}(\mathbf{D}') \quad (57)$$

where $\delta_{\mathbf{D}}(\mathbf{D}')$ can be written as a trajectory MB similar to (54). This explains the last part of Theorem 4 and finishes the proof of Theorem 4.

APPENDIX F

For the particle representation of the multi-trajectory density (28), the cardinality distribution of the set $\mathbf{X}_{\alpha:\gamma}$ of trajectories is given by

$$\Pr(|\mathbf{X}_{\alpha:\gamma}| = n) = \sum_{i=1}^T w^{(i)} \delta_n \left[\left[\mathbf{X}^{(i)} \right] \right], \quad (58)$$

the cardinality distribution of the trajectories born at time step k with $\alpha \leq k \leq \gamma$ is given by

$$\Pr(n \text{ births at time } k) = \sum_{i=1}^T w^{(i)} \delta_n \left[\sum_{(t, x^{1:\nu}) \in \mathbf{X}^{(i)}} \delta_k[t] \right], \quad (59)$$

and the cardinality distribution of the trajectories that die at time step k with $\alpha \leq k \leq \gamma - 1$ is given by

$$\begin{aligned} & \Pr(n \text{ deaths at time } k) \\ &= \sum_{i=1}^T w^{(i)} \delta_n \left[\sum_{(t, x^{1:\nu}) \in \mathbf{X}^{(i)}} \delta_k[t + \nu - 1] \right]. \end{aligned} \quad (60)$$

APPENDIX G

The pseudocode of linear-Gaussian backward simulation for sets of trajectories with PMB filtering densities along with an efficient estimator is given in Algorithm 1. The proposed estimator reports the particle with the highest likelihood accumulated over time from T particles (sets of trajectories) with equal weight $1/T$.

APPENDIX H

In this appendix, we derive the explicit expression of multi-trajectory smoothing density $\pi_{K-1:K|K}(\mathbf{X}_{K-1:K})$ for the one-dimensional example in Section VI-A.

We write $\mathbf{X}_{K-1:K} = \mathbf{Y} \uplus \mathbf{V} \uplus \mathbf{B}$ where the set \mathbf{Y} of trajectories present at both time step $K-1$ and K , the set \mathbf{V} of trajectories only present at time step $K-1$, and the set \mathbf{B} of trajectories only present at time step K . We also write the multi-object filtering density $f_{k|k}(\mathbf{x})$ at time step k as $\delta_{\mathbf{y}_k}(\mathbf{x})$ where $\mathbf{y}_k = \{y_k^1, y_k^2\}$ and $k \in \{1, \dots, K\}$. The predicted multi-trajectory density of $f_{K-1|K-1}(\cdot)$ for each of the events can then be expressed using Theorem 1:

- The two objects existed at both time step $K-1$ and K ,

$$\begin{aligned} & \pi_{K-1:K|K-1}(\mathbf{Y} \uplus \mathbf{B}) \\ &= \delta_{\mathbf{y}_{K-1}}(\tau^{K-1}(\mathbf{Y})) e^{-0.1} \prod_{(K, x^1) \in \mathbf{B}} \lambda^B(x^1) \\ & \quad \times (p^S)^2 \prod_{(K-1, x^{1:2}) \in \mathbf{Y}} \mathcal{N}(x^2; Fx^1, Q). \end{aligned}$$

- Object with state y_{K-1}^1 existed at time step K and object with state y_{K-1}^2 died at time step $K-1$,

$$\begin{aligned} & \pi_{K-1:K|K-1}(\mathbf{Y} \uplus \mathbf{V} \uplus \mathbf{B}) \\ &= \delta_{\mathbf{y}_{K-1}}(\tau^{K-1}(\mathbf{Y} \uplus \mathbf{V})) e^{-0.1} \prod_{(K, x^1) \in \mathbf{B}} \lambda^B(x^1) \\ & \quad \times (1 - p^S) p^S \prod_{(K-1, x^{1:2}) \in \mathbf{Y}} \mathcal{N}(x^2; Fy_{K-1}^1, Q). \end{aligned}$$

- Object with state y_{K-1}^2 existed at time step K and object with state y_{K-1}^1 died at time step $K-1$,

$$\begin{aligned} & \pi_{K-1:K|K-1}(\mathbf{Y} \uplus \mathbf{V} \uplus \mathbf{B}) \\ &= \delta_{\mathbf{y}_{K-1}}(\tau^{K-1}(\mathbf{Y} \uplus \mathbf{V})) e^{-0.1} \prod_{(K, x^1) \in \mathbf{B}} \lambda^B(x^1) \\ & \quad \times (1 - p^S) p^S \prod_{(K-1, x^{1:2}) \in \mathbf{Y}} \mathcal{N}(x^2; Fy_{K-1}^2, Q). \end{aligned}$$

- The two objects died at time step $K-1$,

$$\begin{aligned} & \pi_{K-1:K|K-1}(\mathbf{V} \uplus \mathbf{B}) = \\ & \delta_{\mathbf{y}_{K-1}}(\tau^{K-1}(\mathbf{V})) e^{-0.1} \prod_{(K, x^1) \in \mathbf{B}} \lambda^B(x^1) (1 - p^S)^2. \end{aligned}$$

Algorithm 1 Backward simulation for sets of trajectories with PMB filtering densities

Input: $T, \left\{ \lambda_k^B(\cdot), \lambda_{k|k}^u(\cdot) \right\}_{k=1}^{K-1}, \left\{ \left\{ r_{k|k}^j, p_{k|k}^j(\cdot) \right\}_{j=1}^{n_{k|k}} \right\}_{k=1}^K$.

Output: $\mathbf{X}_{1:K}$

- 1: **for** $\iota = 1, \dots, T$ **do**
- 2: $\mathbf{Y}_{K|K} = \emptyset$;
- 3: $c_\iota = 0$;
- 4: **for** $i = 1, \dots, n_{K|K}$ **do**
- 5: $u \sim \text{Uniform}[0, 1]$;
- 6: **if** $u \leq r_{K|K}^i$ **then**
- 7: $\mathbf{y}_{K|K} = \mathbf{y}_{K|K} \cup \{x_{K|K}^i\}$;
- 8: **end if**
- 9: **end for**
- 10: $\mathbf{Y}_{K:K} = \{Y = (1, y^1) : y^1 \in \mathbf{y}_{K|K}\}$
- 11: **for** $k = K-1, \dots, 1$ **do**
- 12: Separate $\mathbf{Y}_{k+1:K}$ into $\{Y^j\}_{j=1}^m$ and $\{Y^j\}_{j=m+1}^{n_{k+1:K}}$ as described in Theorem 4;
- 13: **for** $i = 1, \dots, n_{k|k}$ **do**
- 14: **for** $j = 1, \dots, m$ **do**
- 15: **if** $\text{SMD}(y_j^i; x_{k|k}^i, P_{k|k}^i) < \Gamma_g$ **then**
- 16: Compute $W_1^{(j,i)}$ using (36b);
- 17: **else**
- 18: $W_1^{(j,i)} = 0$;
- 19: **end if**
- 20: **end for**
- 21: **end for**
- 22: Compute W_2 (36c);
- 23: $C = -\log[W_1 \quad W_2]$;
- 24: Run Murty's algorithm on C to obtain the M best global hypotheses with highest weight $a_{k:K|K}^* = \{a_1, \dots, a_M\}$;
- 25: Compute $[\hat{w}_{a_1}, \dots, \hat{w}_{a_M}]$ using (18) and normalise them to obtain $[w_{a_1}, \dots, w_{a_M}]$;
- 26: $a \sim \text{Categorical}([w_{a_1}, \dots, w_{a_M}])$;
- 27: $c_\iota = c_\iota + \log(\hat{w}_a)$;
- 28: **end for**
- 29: $\mathbf{Y}_{k:K} = \{Y^j\}_{j=m+1}^{n_{k+1:K}}$;
- 30: **for** $i = 1, \dots, n_{k|k}$ **do**
- 31: **if** $a^i = 1$ **then**
- 32: $u \sim \text{Uniform}[0, 1]$;
- 33: **if** $u \leq r_{k:K|K}^{i,1}$ (32b) **then**
- 34: Sample $Y \sim p_{k:K|K}^{i,1}(\cdot)$ using (32c);
- 35: $\mathbf{Y}_{k:K} = \mathbf{Y}_{k:K} \cup \{Y\}$;
- 36: **end if**
- 37: **else**
- 38: Sample $Y \sim p_{k:K|K}^{i,a^i}(\cdot)$ using (33c);
- 39: $\mathbf{Y}_{k:K} = \mathbf{Y}_{k:K} \cup \{Y\}$;
- 40: **end if**
- 41: **end for**
- 42: **for** $i = n_{k|k} + 1, \dots, n_{k|k} + m$ **do**
- 43: **if** $a^i = 2$ **then**
- 44: Sample $Y \sim p_{k:K|K}^{i,2}(\cdot)$ using (34c);
- 45: $\mathbf{Y}_{k:K} = \mathbf{Y}_{k:K} \cup \{Y\}$;
- 46: **end if**
- 47: **end for**
- 48: $\mathbf{X}_{1:K}^{(\iota)} = \mathbf{Y}_{1:K}$;
- 49: **end for**
- 50: $\iota^* = \arg \max_\iota c_\iota$;
- 51: $\mathbf{X}_{1:K} = \mathbf{X}_{1:K}^{(\iota^*)}$;

When taking the product of the two multi-trajectory densities $\pi_{K-1:K|K}(\mathbf{X}_{K-1:K})$ and $\pi_{K:K|K}(\mathbf{X}_{K|K})$, a multi-trajectory Dirac delta with two components, each of the first three events has two different cases due to the unknown correspondence between $\tau^{K-1}(\mathbf{Y})$ and \mathbf{y}_K . Therefore, the multi-trajectory density $\pi_{K-1:K|K}(\mathbf{X}_{K-1:K})$ has a mixture representation of 7 components.