Information-Theoretic Joint Probabilistic Data Association Filter

Shaoming He, Hyo-Sang Shin, and Antonios Tsourdos

***Abstract*—This article proposes a novel information-theoretic joint probabilistic data association filter for tracking unknown number of targets. The proposed information-theoretic joint probabilistic data association algorithm is obtained by the minimization of a weighted reverse Kullback–Leibler divergence to approximate the posterior Gaussian mixture probability density function. Theoretical analysis of mean performance and error covariance performance with ideal detection probability is presented to provide insights of the proposed approach. Extensive empirical simulations are undertaken to validate the performance of the proposed multitarget tracking algorithm.**

***Index Terms*—Information-theoretic approach, joint probabilistic data association, multiple target tracking.**

I. Introduction

Multitarget tracking (MTT) is an important and fundamental technology in many engineering applications, including impact point prediction [1], airborne surveillance [2], space situation awareness [3], autonomous navigation [4], and computer vision [5]. The main objective of MTT algorithm is to find tracks from noisy unlabeled measurements. The issue of MTT is the uncertain source of the measurements: that is, the mappings between the targets and the measurements are unknown. Therefore, MTT algorithms usually require a data association process to work out which measurement is originated from which target.

The well-established multiple hypothesis tracking (MHT) filter [6], [7] maintains all possible data association hypotheses in a decision-making tree with accumulated measurements. This tracking filter finds the global association hypothesis with the highest probability by a delayed decision logic for estimation update. Although MHT is mathematically Bayesian optimal, the exact solution is known as a NP-hard problem and, thus, is computationally intractable for large-scale scenarios. Another frequently used data association algorithm, known as joint probabilistic data association (JPDA), is a suboptimal approach to the Bayesian filter [8], [9]. This algorithm makes soft decisions on measurement-to-target associations, allowing for the possibility that a measurement may have originated from a number of candidate targets. Compared to MHT, JPDA can achieve reasonable results with lower computational burden.

Due to the nature of the soft decision, the posterior distribution provided by JPDA is a Gaussian mixture probability density function (PDF) with each Gaussian term representing one possible measurement-to-target association. Since the propagation of the Gaussian mixture is computationally intractable in the implementation, JPDA approximates the posterior Gaussian mixture by a single Gaussian using simple moment-preserving approach. Clearly, this simple Gaussian mixture reduction method can be utilized to approximate the posterior as long as the Gaussian terms are well-spaced. If two Gaussian terms are close enough, e.g., targets are moving closely, the posterior Gaussian mixture will be highly bimodal (or multimodal), and therefore, the single Gaussian approximation made by JPDA is not accurate enough to represent the real posterior PDF [10]–[12]. This fact means that the JPDA filter tends to ignore the interactions between those close Gaussian terms and the estimation performance will degrade. Instead of enumerating all possible joint association events, the exact nearest neighbor JPDA (ENNJPDA) [10] only picks up the joint association event with the highest probability for marginalization. This drastic pruning strategy explicitly avoids the multimodal issue by utilizing only one joint association event, but it causes significant information loss as only one joint association event is used for marginalization. This implies that the ENNJPDA is sensitive to the clutter rate. An adjusted version of the JPDA filter, which is called the Set JPDA filter [11], is proved to be an effective way to preserve the multimodality of the posterior Gaussian mixture PDF. However, this algorithm assumes the number of targets is known *a priori* and is unable to preserve the target identity. The multimodality problem of JPDA is directly avoided in [13] via a novel measurement driven approach. The measurement-driven JPDA, however, is sensitive to measurement noise and cannot provide the information on target label. By Bayesian variational approach, the authors in [12] proposed another JPDA to approximate the posterior PDF, but this approach requires an iterative optimization loop for the implementation and, therefore, is computationally expensive.

Motivated by the aforementioned observations, the authors aim to propose a MTT algorithm that can efficiently improve the tracking performance of the original JPDA. More specifically, an information-theoretic approach is utilized to approximate the posterior Gaussian mixture PDF. This approximation is obtained by the minimization of an information-theoretic metric, i.e., a weighted reverse Kullback–Leibler divergence (KLD), and the resultant formula shows similar structure as generalized covariance intersection (GCI) in sensor fusion. The motivation behind using KLD as a metric is that it quantifies the similarity between two PDFs. This enables us to find a best approximation of a specific PDF in an information-theoretic way. Unlike in ENNJPDA, leveraging the weighted KLD enables the resultant PDF to approach the mode with the most probable association pair without completely pruning other information. Therefore, the proposed approach could alleviate the multimodal issue while providing a certain level of robustness against clutters. Although the basic reverse KLD-based Gaussian mixture reduction is available in the literature, see [14], for example, no open publication utilizes this technique to refine the update rules of JPDA. The uniqueness of this article is in choosing the weights for the mixture as the marginal association probabilities over the basic reverse KLD mixture model. Another benefit of using a weighted KLD as the cost function is that it enables a simple closed-loop solution, which largely resembles the concept of GCI and, therefore, provides

Manuscript received September 2, 2019; revised January 10, 2020; accepted April 17, 2020. Date of publication May 28, 2020; date of current version February 26, 2021. Recommended by Associate Editor Sean B. Andersson. *(Corresponding author: Hyo-Sang Shin.)*

The authors are with the School of Aerospace, Transport and Manufacturing, Cranfield University, Cranfield MK43 0AL, U.K. (e-mail: shaoming.he@cranfield.ac.uk; h.shin@cranfield.ac.uk; a.tsourdos@cranfield.ac.uk).

Color versions of one or more of the figures in this article are available online at https://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TAC.2020.2989766

robustness against the cross correlations between two Gaussian terms. In the case of ideal detection probability, the performance bounds of the error covariance and the mean performance is theoretically analyzed to support the validity of the proposed approach. Extensive empirical simulations are undertaken to validate the performance of the proposed MTT algorithm for scenarios with closely moving targets.

II. Backgrounds and Preliminaries

A. System Model

The set of target states and measurements received at scan k are, respectively, defined as

$$X_k = \left\{x_k^1, \ldots, x_k^{N_k}\right\}, \quad Z_k = \left\{z_k^0, z_k^1, \ldots, z_k^{M_k}\right\} \tag{1}$$

where N_k denotes the number of targets at scan k, x_k^i the ith target at scan k, M_k the number of measurements received at scan k, $z_k^j (j \neq 0)$ the jth measurement received at scan k, z_k^0 the dummy measurement for convenient representation of miss detection. The temporal evolution of each target is independent of others and follows a Markov transition model $p(x_k^i | x_{k-1}^i)$, which is determined by a linear system as

$$x_k^i = F_{k-1}^i x_{k-1}^i + w_{k-1}^i, \quad z_k^i = H_k^i x_k^i + v_k^i \tag{2}$$

where $x_k^i \in \mathbb{R}^n$ and $z_k^i \in \mathbb{R}^m$ denote the system state and the corresponding measurement at time step k, respectively. The notations F_{k-1}^i and H_k^i correspond to the system matrix and observation matrix, respectively. The process noise w_{k-1}^i and measurement noise v_k^i are assumed to be independent. These two signals are modeled as a zero-mean Gaussian process with covariances Q_{k-1}^i and R_k^i, respectively.

For convenience, we make the following general assumptions, which are widely accepted in MTT problems.

Assumption 1: Each target can generate at most one measurement and each measurement can originate from at most one target. Each target-generated measurement is independent of each other and is detected with probability P_D with measurement likelihood $p(z|x)$.

Assumption 2: The clutter distribution is assumed to be unknown *a priori* and is, thus, considered as Poisson distribution. Clutters or false alarms are modeled by a local Poisson point process (PPP) with spatial intensity $\lambda_{\text{FA}}(z)$.

In MTT problem, the number of targets is usually unknown due to target birth/death and, therefore, track management, i.e., target initialization and deletion, is required for MTT algorithms. For this reason, we utilize a PPP model with intensity $\lambda_B(x)$, similar to [13], for target birth and perform track confirmation/deletion by the existence probability thresholding approach [15]. Within this framework, measurements that originate from extraneous source, either new targets or false alarms, can be modeled by a PPP with intensity $\lambda_E = \lambda_B(x_b) p(z|x_b) P_D + \lambda_{FA}(z)$, where x_b denotes the possible state vector of new birth targets. Let $\mathcal{X}_k = \{\chi_k^1, \ldots, \chi_k^{N_k}\}$ denote the event of target existence at scan k, where χ_k^i represents the event of the ith target existence. Then, the prediction or time evolution of χ_k^i can be formulated by the Markov Chain One model [15] as $p(\chi_k^i | Z_{k-1}) = P_S p(\chi_{k-1}^i | Z_{k-1})$, where P_S denotes the target survival probability.

B. Problem Formulation

JPDA aims to calculate the marginalized association probability for estimation update by enumerating all possible joint hypotheses. A feasible joint hypothesis in JPDA is defined as one possible measurement-to-target mapping such that: (1) each measurement (except for the dummy one) is assigned to at most one target; (2) each target is uniquely assigned to a measurement. Let $\Theta_k = \{\theta_k^i\}$, $i \in \{1, 2, \ldots, N_{k-1} + M_k\}$, denote the joint association vector. For each pre-existed target $i \in \{1, 2, \ldots, N_{k-1}\}$, define $\theta_k^i \in \{0, 1, \ldots, M_k\}$ as the association hypothesis, and let $\theta_k^i = j$ denote the event that the jth measurement is originated from the ith target with $\theta_k^i = 0$ representing a miss-detection event. We create a new track for each measurement $j \in \{1, 2, \ldots, M_k\}$ at scan k, and the association events for these new targets are defined by $\theta_k^{N_{k-1}+j} \in \{N_{k-1} + 1, \ldots, N_{k-1} + M_k\}$. That is, if target $N_{k-1} + j$ is associated with the jth measurement, then $\theta_k^{N_{k-1}+j} = N_{k-1} + j$. Under the assumption that each single association event is independent, the posterior of each target can be modeled by a mixture term as

$$p\left(x_k^i \middle| \chi_k^i, Z_k\right) = \sum_{\theta_k^i} p\left(x_k^i \middle| \theta_k^i, \chi_k^i, Z_k\right) p\left(\theta_k^i \middle| \chi_k^i, Z_k\right). \tag{3}$$

Generally, the number of mixture terms increases exponentially due to the data association uncertainty. Since every mixture term requires a Kalman filter for its prorogation, it is necessary to control the number of mixture terms from the computational burden standpoint. The standard JPDA simply approximates the posterior mixture PDF by a single distribution through the moment-preserving method. More specifically, the state correction $x_{k|k}^i$ of the ith target and its corresponding covariance $P_{k|k}^i$ are obtained as

$$\begin{aligned} x_{k|k}^i &= \sum_{j=0}^{M_k} \beta_j^i x_{k|k}^{i,j} \\ P_{k|k}^i &= \sum_{j=0}^{M_k} \beta_j^i \left\{ P_{k|k}^{i,j} + \left(x_{k|k}^{i,j} - x_{k|k}^i\right)\left(x_{k|k}^{i,j} - x_{k|k}^i\right)^T \right\} \end{aligned} \tag{4}$$

where $\{x_{k|k}^{i,j}, P_{k|k}^{i,j}\}$ denotes the estimate by associating the jth measurement to the ith target and $\beta_j^i = p(\theta_k^i = j | \chi_k^i, Z_k)$ is the existence-conditioned marginal association probability that the jth measurement is associated with the ith target.

As stated in [16], the moment-preserving approximation, which simply merges all Gaussian mixtures, is accurate enough provided that the distance between Gaussian terms is far enough. If two targets are not well-spaced or in a highly cluttered environment, the resulting Gaussian mixture exhibits multimodality, and thus, this approximation may destroy valuable information. Motivated by these observations, this article aims to suggest a MTT algorithm that approximates the Gaussian mixture PDF in a more accurate way to improve the performance of the JPDA filter through an information-theoretic approach.

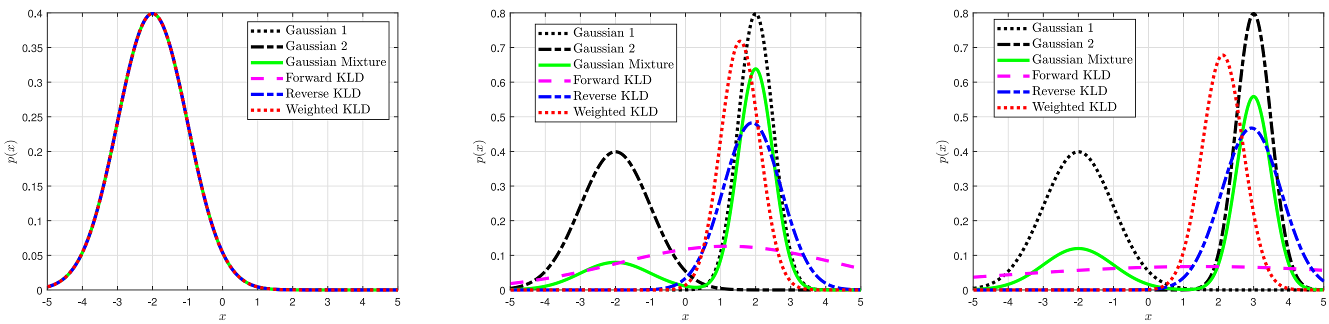
III. Information-Theoretic JPDA

A. Mixture Approximation by Information-Theoretic Approach

Instead of simple moment-matching, we will propose an information-theoretic approach for Gaussian mixture reduction in JPDA. In general, there are many ways to quantify the proximity of two PDFs. A detailed comparison of different statistical and information-theoretic measures in mixture reduction has been presented in [14] and [17]. The results demonstrate that the utilization of the information-theoretic KLD could be a competitive alternative. The KLD or cross entropy is defined as

$$\mathcal{D}_{\text{KL}}(p \| q) = \int p(x) \ln \frac{p(x)}{q(x)} dx \tag{5}$$

which reveals that $\mathcal{D}_{\text{KL}}(p \| q) \geq 0$ with equality if and only if $p(x) = q(x)$. As the KLD quantifies how close a PDF is to a candidate PDF,



(a) (b) (c)

Fig. 1. Comparison of different Gaussian mixture approximations. (a) $p(x) = 0.5\mathcal{N}(x; -2, 1) + 0.5\mathcal{N}(x; -2, 1)$. (b) $p(x) = 0.8\mathcal{N}(x; 2, 0.5) + 0.2\mathcal{N}(x; -2, 1)$. (c) $p(x) = 0.3\mathcal{N}(x; -2, 1) + 0.7\mathcal{N}(x; 3, 0.5)$.

one can find a best approximation of a specific PDF in the sense of minimization of KLD. Let $p(x)$ be the original PDF and $q(x)$ the candidate PDF that approximates the original one. In information theory, there are two different kinds of KLDs, forward KLD and reverse KLD, that we can seek to minimize to obtain our solution $q(x)$. These two different KLDs are defined as [14]

$$\text{Forward KLD:} \quad \mathcal{D}_{\text{KL}}\left(p \,\|q\right) = \int p(x) \ln \frac{p(x)}{q(x)} dx$$
$$\text{Reverse KLD:} \quad \mathcal{D}_{\text{KL}}\left(q \,\|p\right) = \int q(x) \ln \frac{q(x)}{p(x)} dx. \tag{6}$$

The solution of $\min_q \mathcal{D}(p\|q)$ is known as moment projection, which finds the mean of $p(x)$ and overestimate the support of $p(x)$. In [18], the authors proposed a Gaussian mixture reduction method on the basis of the forward KLD, which was proved to only consider the merging operation for mixture reduction [14]. As a comparison, the minimization of the reverse KLD results in a mode selection approximation [14]. Although the mode seeking property of the reverse KLD can explicitly avoid the issue of multimodality, it does not permit an analytic solution. This means that the utilization of reverse KLD in JPDA update requires computationally expensive numerical optimizations, which might not be suitable for the ever-increasing low-cost sensors [19]. To address this problem, we propose to minimize an approximated reverse KLD, i.e.,

$$\arg\min_{p^i} \sum_{j=0}^{M_k} \beta_j^i \mathcal{D}_{\text{KL}}\left(p^i \,\|p_j^i\right) \tag{7}$$

where $p_j^i(x^i) = \mathcal{N}(x^i; x_{k|k}^{i,j}, P_{k|k}^{i,j})$ denotes the PDF representing the relationship between the ith target and the jth measurement, p^i the candidate approximated posterior PDF of the ith target.

Since the marginal association probability β_j^i quantifies the confidence level that the jth measurement comes from the ith target, β_j^i is leveraged as the weight/penalty for the difference between the candidate PDF p_i and the single association Gaussian term $p_j^i(x^i)$. This implies that the proposed cost function also considers matching with the strongest Gaussian mode, i.e., with the highest marginal association probability, to resolve the multimodality issue when targets are moving very closely. The mathematical relationship between the original reverse KLD and the proposed cost function is established in the following proposition.

Proposition 1: The reverse KLD provides the lower bound of the proposed cost function, i.e.,

$$\mathcal{D}_{\text{KL}}\left(p^i \,\middle\|\, \sum_{j=0}^{M_k} \beta_j^i p_j^i\right) \leq \sum_{j=0}^{M_k} \beta_j^i \mathcal{D}_{\text{KL}}\left(p^i \,\|p_j^i\right) \tag{8}$$

where the equality holds if and only if $p_{j'}^i = p_j^i, \forall j' \neq j$.

Proof: Please refer to Appendix A. ■

Remark 1: Proposition 1 indicates that the proposed cost function reduces to the reverse KLD once all Gaussian terms have the same mean and covariance. Fig. 1 provides a comparison example of different Gaussian mixture approximation approaches, where the solution of minimizing the reverse KLD is obtained via numerical algorithms. From Fig. 1(a), it can be clearly noted that minimizing the reverse KLD and the proposed weighted KLD generate the same solution if all Gaussian terms have the same distribution. Comparing to the moment matching, i.e., minimizing the forward KLD, optimizing the reverse KLD provides the opportunity of mode selection and, therefore, can directly resolve the issue of multimodality in JPDA. As an approximation of the reverse KLD, the proposed cost function provides the solution with approximate mode selection, as confirmed by Fig. 1.

Proposition 2: The closed-form solution of optimization problem (7) is given by a Gaussian distribution

$$p_*^i = \mathcal{N}\left(x^i; x_{k|k}^i, P_{k|k}^i\right) \tag{9}$$

with

$$x_{k|k}^i = P_{k|k}^i \sum_{j=0}^{M_k} \beta_j^i \left(P_{k|k}^{i,j}\right)^{-1} x_{k|k}^{i,j}$$
$$P_{k|k}^i = \left[\sum_{j=0}^{M_k} \beta_j^i \left(P_{k|k}^{i,j}\right)^{-1}\right]^{-1}. \tag{10}$$

Proof: Please refer to Appendix B. ■

Remark 2: Proposition 2 reveals that the optimal solution is the weighted geometric mean of all PDFs, which coincides with the GCI in information fusion [20]. This means that the proposed algorithm holds similar property as GCI. That is, the proposed method is robust against the unknown correlations among the information sources.

Remark 3: By using the computational complexity of the related matrix manipulations, the computational complexity comparisons of the proposed Gaussian mixture approximation and the simple moment-preserving approach of the original JPDA are presented in Table I .

TABLE I
COMPUTATIONAL COMPLEXITY COMPARISON

Algorithm	Computation term	Computational complexity
Moment-preserving	$x^i_{k\|k} = \sum_{j=0}^{M_k} \beta^i_j x^{i,j}_{k\|k}$	$\mathcal{O}\left((M_k+1)\, n\right)$
	$P^i_{k\|k} = \sum_{j=0}^{M_k} \beta^i_j \left\{ P^{i,j}_{k\|k} + \left(x^{i,j}_{k\|k} - x^i_{k\|k}\right)\left(x^{i,j}_{k\|k} - x^i_{k\|k}\right)^T \right\}$	$\mathcal{O}\left((M_k+1)\left(3n^2+n\right)\right)$
Proposed approach	$x^i_{k\|k} = P^i_{k\|k} \sum_{j=0}^{M_k} \beta^i_j \left(P^{i,j}_{k\|k}\right)^{-1} x^{i,j}_{k\|k}$	$\mathcal{O}\left((M_k+1)\left(n^3+2n^2\right)+n^2\right)$
	$P^i_{k\|k} = \left[\sum_{j=0}^{M_k} \beta^i_j \left(P^{i,j}_{k\|k}\right)^{-1}\right]^{-1}$	$\mathcal{O}\left((M_k+1)\left(n^3+n^2\right)+n^3\right)$

From this table, one can note that the proposed approach has higher computational complexity than the simple moment-matching method due to the inversion operation of $P^{i,j}_{k|k}$. However, the computational complexities of both algorithms are scalable with respect to the number of measurements.

B. Performance Analysis

In this section, the performance of the proposed mixture approximation (9) is analyzed to provide better insights of its properties. Intuitively, continuous miss detection results in significant information loss, leading to the divergence of the estimation filter. Therefore, the performance is analyzed under condition $P_D = 1$ in this section, while the performance of the proposed algorithm in a more realistic scenario will be empirically studied in the simulation parts.

Theorem 1: (Mean Performance) Under the condition $P_D = 1$, the proposed approach (9) yields unbiased estimation for linear system (2) if $\mathbb{E}[x^{i,j}_{k|k}] = x^i_k$.

Proof: Define $Y^i_{k|k} = (P^i_{k|k})^{-1}$ and $y^i_{k|k} = (P^i_{k|k})^{-1} x^i_{k|k}$. Then, (9) can be rewritten as

$$y^i_{k|k} = \sum_{j=0}^{M_k} \beta^i_j y^{i,j}_{k|k}, \quad Y^i_{k|k} = \sum_{j=0}^{M_k} \beta^i_j Y^{i,j}_{k|k} \tag{11}$$

which reveals that the proposed mixture reduction algorithm can be viewed as a weighted sum of information terms.

The remaining proof is given by mathematical induction. Consider the case that the ith target has two valid measurements, e.g., $j = 1, 2$. Then, (11) reduces to

$$\begin{aligned} y^i_{k|k} &= \beta^i_1 y^{i,1}_{k|k} + \left(1-\beta^i_1\right) y^{i,2}_{k|k} \\ Y^i_{k|k} &= \beta^i_1 Y^{i,1}_{k|k} + \left(1-\beta^i_1\right) Y^{i,2}_{k|k}. \end{aligned} \tag{12}$$

Applying the matrix inversion lemma, we have

$$\left(Y^i_{k|k}\right)^{-1}\left(\beta^i_1 Y^{i,1}_{k|k}\right) = I - B \tag{13}$$

where

$$B \triangleq \left(\beta^i_1 Y^{i,1}_{k|k}\right)^{-1} A^{-1}, A \triangleq \left(\beta^i_1 Y^{i,1}_{k|k}\right)^{-1} + \left[\left(1-\beta^i_1\right) Y^{i,2}_{k|k}\right]^{-1}. \tag{14}$$

Through simple manipulations, we have

$$\left(1-\beta^i_1\right) Y^{i,2}_{k|k}\left(\beta^i_1 Y^{i,1}_{k|k}\right)^{-1} + I = \left(1-\beta^i_1\right) Y^{i,2}_{k|k} A \tag{15}$$

$$lI + \left(1-\beta^i_1\right) Y^{i,2}_{k|k}\left(\beta^i_1 Y^{i,1}_{k|k}\right)^{-1} = Y^i_{k|k}\left(\beta^i_1 Y^{i,1}_{k|k}\right)^{-1}. \tag{16}$$

Combining (15) and (16) yields

$$\left(Y^i_{k|k}\right)^{-1}\left(1-\beta^i_1\right) Y^{i,2}_{k|k} = \left(\beta^i_1 Y^{i,1}_{k|k}\right)^{-1} A^{-1} = B. \tag{17}$$

Based on (13) and (17), it is straightforward to evaluate the expectation $\mathbb{E}[x^i_{k|k}]$ as

$$\begin{aligned} \mathbb{E}\left[x^i_{k|k}\right] &= \mathbb{E}\left[\left(Y^i_{k|k}\right)^{-1} y^i_{k|k}\right] \\ &= \mathbb{E}\left[\left(Y^i_{k|k}\right)^{-1}\left(\beta^i_1 Y^{i,1}_{k|k} x^{i,1}_{k|k} + \left(1-\beta^i_1\right) Y^{i,2}_{k|k} x^{i,2}_{k|k}\right)\right] \\ &= \mathbb{E}\left[(I-B)\, x^{i,1}_{k|k} + B x^{i,2}_{k|k}\right] \\ &= x^i_k. \end{aligned} \tag{18}$$

At the inductive step, assume that the ith target has $L-1$ valid measurements, e.g., $j = 1, 2, \ldots, L-1$ and the proposed Gaussian approximation approach yields unbiased estimation

$$y^i_{k|k} = \sum_{j=1}^{L-1} \beta^i_j y^{i,j}_{k|k}, \quad Y^i_{k|k} = \sum_{j=1}^{L-1} \beta^i_j Y^{i,j}_{k|k}. \tag{19}$$

Following similar procedures, shown in (12)–(18), it is easy to verify that, if the ith target has L valid measurements, e.g., $j = 1, 2, \ldots, L$, the proposed Gaussian approximation approach yields unbiased estimation. Then, the proof is completed by mathematical induction. ■

Remark 4: Notice that the expectation $\mathbb{E}[x^i_{k|k}]$ is evaluated based on the posterior $p(x^i_k|\chi^i_k, Z_k)$, which is a function of x^i_k given current measurements Z_k. Since data association only depends on the target dynamics model and current measurements Z_k, the marginal probability β^i_j is determined once the association process is finished and therefore β^i_j is considered as a known constant in the analysis of Theorem 1.

Remark 5: The establishment of Theorem 1 requires the assumption that every hypothesis leads an unbiased estimate. Notice that each single association hypothesis is updated by the classical linear Kalman filter, which provides unbiased estimate provided that the noises are Gaussian. For this reason, we utilize the unbiased assumption of each hypothesis in the analysis.

Theorem 1 reveals that, if the original estimator for single association pair update is unbiased, the proposed algorithm yields unbiased estimation. Now, let us investigate the error covariance performance, which quantifies the estimation accuracy, of the propose information-theoretic approach. Before providing the results, a useful lemma regarding the boundedness of the Kalman filter [21] is briefly reviewed first. Consider the following linear system:

$$x_k = F_{k-1} x_{k-1} + w_{k-1}, \quad z_k = H_k x_k + v_k \tag{20}$$

where $x_k \in \mathbb{R}^n$ and $z_k \in \mathbb{R}^m$ denote the system state and the corresponding measurement at time step k, respectively. The notations F_{k-1} and H_k correspond to the system matrix and observation matrix, respectively. The signals w_{k-1} and v_k are process noise and measurement noise, which are assumed to be the zero-mean Gaussian process with covariances Q_{k-1} and R_k, respectively.

Definition 1: System (20) is uniformly controllable and observable if there exist a positive integer k_0 and positive constants $\kappa_1 \le \kappa_2 < \infty$, $\kappa_3 \le \kappa_4 < \infty$ such that $\kappa_1 I \le \sum_{i=k-k_0+1}^{k} \Phi_{k,i} Q_i \Phi_{k,i}^T \le \kappa_2 I$ and $\kappa_3 I \le \sum_{i=k-k_0}^{k} \Phi_{i,k}^T H_i^T R_i^{-1} H_i \Phi_{i,k} \le \kappa_4 I$ with

$$\Phi_{k,l} = \begin{cases} F_{k-1}F_{k-2}\cdots F_l, & k > l \\ I, & k = l \end{cases}, \quad \Phi_{i,k} = \Phi_{k,i}^{-1}. \tag{21}$$

By applying Kalman filter to system (20), we can have the following result [21].

Lemma 1: Suppose system (20) is uniformly controllable and uniformly observable. If the initial error covariance $P_{0|0}$ is positive, then there exists a positive integer k_0 such that the error covariance provided by the Kalman filter is uniformly bounded from below and above for all $k \ge k_0$ as $\underline{P} \le P_{k|k} \le \overline{P}$, where $\underline{P}$ and $\overline{P}$ are positive matrices.

Theorem 2: (Error Covariance Performance) Suppose $P_D = 1$ and the linear system (2) is uniformly controllable and uniformly observable. If the initial error covariance $P_{0|0}^{i,j} > 0$, then the estimation error given by (9) is asymptotically bounded in a mean square sense and the corresponding error covariance is uniformly bounded from below and above as $\underline{P} \le P_{k|k}^i \le \overline{P}$, where $\underline{P}$ and $\overline{P}$ are positive matrices.

Proof: According to JPDA, the error covariance update of each association pair is obtained by the linear Kalman filter as

$$\begin{aligned}\left(P_{k|k}^{i,j}\right)^{-1} &= \left(P_{k|k-1}^{i,j}\right)^{-1} + \left(H_k^i\right)^T \left(R_k^i\right)^{-1} H_k^i \\ &= \left\{F_{k-1}^i \left[\sum_{j=1}^{M_k} \beta_j^i \left(P_{k-1|k-1}^{i,j}\right)^{-1}\right]^{-1} \left(F_{k-1}^i\right)^T + Q_{k-1}^i\right\}^{-1} \\ &\quad + \left(H_k^i\right)^T \left(R_k^i\right)^{-1} H_k^i. \end{aligned} \tag{22}$$

Note that the previous equation is derived based on the information form of the Kalman filter. According to Lemma 1, the hypothesis-conditioned error covariance is uniformly bounded from below and from above as $\underline{P}_i \le P_{k|k}^{i,j} \le \overline{P}_i$, where $\underline{P}_i$ and $\overline{P}_i$ are positive matrices. Based on this fact, at each scan, we can choose the upper and lower bounds of $P_{k|k}^{i,j}$, i.e., $P_{k|k}^{i,U}$ and $P_{k|k}^{i,L}$, such that

$$P_{k|k}^{i,U} - P_{k|k}^{i,j} \ge 0, \quad P_{k|k}^{i,L} - P_{k|k}^{i,j} \le 0 \quad \forall j \in [M_k] \tag{23}$$

where $[M_k] \triangleq \{1, 2, \ldots, M_k\}$. Then, one can imply that

$$\begin{aligned} \sum_{j=1}^{M_k} \beta_j^i \left(P_{k|k}^{i,j}\right)^{-1} &\le \sum_{j=1}^{M_k} \beta_j^i \left(P_{k|k}^{i,L}\right)^{-1} = \left(P_{k|k}^{i,L}\right)^{-1} \\ \sum_{j=1}^{M_k} \beta_j^i \left(P_{k|k}^{i,j}\right)^{-1} &\ge \sum_{j=1}^{M_k} \beta_j^i \left(P_{k|k}^{i,U}\right)^{-1} = \left(P_{k|k}^{i,U}\right)^{-1}. \end{aligned} \tag{24}$$

Substituting (24) into (22) yields

$$\begin{aligned} \left(P_{k|k}^{i,j}\right)^{-1} &\le \left(F_{k-1}^i P_{k-1|k-1}^{i,L} \left(F_{k-1}^i\right)^T + Q_{k-1}^i\right)^{-1} \\ &\quad + \left(H_k^i\right)^T \left(R_k^i\right)^{-1} H_k^i \\ \left(P_{k|k}^{i,j}\right)^{-1} &\ge \left(F_{k-1}^i P_{k-1|k-1}^{i,U} \left(F_{k-1}^i\right)^T + Q_{k-1}^i\right)^{-1} \\ &\quad + \left(H_k^i\right)^T \left(R_k^i\right)^{-1} H_k^i. \end{aligned} \tag{25}$$

Note that the right-hand sides of (25) can be considered as two standard Kalman filter error covariance updates as

$$\begin{aligned} P_{k|k}^{i,j} &\ge \Psi_k\left(\Psi_{k-1}\cdots\left(\Psi_1\left(P_{0|0}^{i,L}\right)\cdots\right)\right) \\ P_{k|k}^{i,j} &\le \Psi_k\left(\Psi_{k-1}\cdots\left(\Psi_1\left(P_{0|0}^{i,U}\right)\cdots\right)\right) \end{aligned} \tag{26}$$

Algorithm 1: Information-Theoretic JPDA Filter.

1: Predict the existing targets using target dynamics
2: Update each track with all possible measurements by Kalman filter
3: Calculate the marginal probability $p(\theta_k^i|Z_k)$
4: Calculate the posterior existence probability $p(\chi_k^i|Z_k)$
5: Calculate the existence-conditioned marginal probability $p(\theta_k^i|\chi_k^i, Z_k)$
6: Use Proposition 2 to update N_{k-1} existing targets
7: Generate M_k new targets (one for each measurement)
8: Perform track confirmation and deletion based using the existence probability $p(\chi_k^i|Z_k)$

where the operator $\Psi_k(\cdot)$ is defined as

$$\begin{aligned} \Psi_k(P) \triangleq \Bigg[&\left(F_{k-1}^i P \left(F_{k-1}^i\right)^T + Q_{k-1}^i\right)^{-1} \\ &+ \left(H_k^i\right)^T \left(R_k^i\right)^{-1} H_k^i\Bigg]^{-1}. \end{aligned} \tag{27}$$

Consequently, the estimation error given by (9) is asymptotically bounded in a mean square sense and the corresponding error covariance is uniformly bounded from below and above, following Lemma 1. ■

IV. Implementation of Proposed JPDA

This section presents the details of the implementation of the proposed information-theoretic JPDA filter. The implementation details of the proposed information-theoretic JPDA is summarized in Algorithm 1.

A. Implementation Details

Theoretically, the marginal association probability $p(\theta_k^i|Z_k)$ can be obtained by finding all the feasible joint association hypotheses as

$$p\left(\theta_k^i = j\,|Z_k\right) = \sum\nolimits_{\theta_k^i(\in\Theta_k)=j} p\left(\Theta_k\,|Z_k\right) \tag{28}$$

where the posterior distribution of the joint association hypothesis $p(\Theta_k|Z_k)$ is given by $p(\Theta_k|Z_k) \propto f_m \times f_d \times f_n$, where f_m, f_d, and f_n refer to the terms related to miss-detection, target detection, and new target, respectively, as

$$\begin{aligned} f_m &\propto \left[\prod_{i\in[N_{k-1}],\theta_k^i=0} 1 - P_D p\left(\chi_k^i\,|Z_{k-1}\right)\right] \\ f_d &\propto \left[\prod_{i\in[N_{k-1}],\theta_k^i=j} P_D p\left(\chi_k^i\,|Z_{k-1}\right) p\left(z_k^j\,\middle|x_{k|k-1}^i\right)\right] \\ f_n &\propto \left[\prod_{\theta_k^{N_{k-1}+j}=N_{k-1}+j} \sum_{s=1}^{N_b} P_D \lambda_B\left(x_b^s\right) p\left(z_k^j\,|x_b^s\right) + \lambda_{\text{FA}}\left(z_k^j\right)\right] \end{aligned} \tag{29}$$

where x_b^s denotes sth possible state vector of new birth targets and N_b represents the total number of possible initializations for new targets.

After finding $p(\theta_k^i|Z_k)$, the joint probability $p(\theta_k^i, \chi_k^i|Z_k)$ can be calculated using the chain rule of probability as

$$p\left(\theta_k^i, \chi_k^i\,|Z_k\right) = p\left(\chi_k^i\,\middle|\theta_k^i, Z_k\right) p\left(\theta_k^i\,|Z_k\right) \tag{30}$$

where the hypothesis-conditioned existence probability $p(\chi_k^i|\theta_k^i, Z_k)$ can be easily obtained as

$$p\left(\chi_k^i\left|\theta_k^i, Z_k\right.\right) \propto \begin{cases} \frac{p\left(\chi_k^i|Z_{k-1}\right)(1-P_D)}{1-p\left(\chi_k^i|Z_{k-1}\right)+p\left(\chi_k^i|Z_{k-1}\right)(1-P_D)}, \theta_k^i = 0 \\ 1, \qquad\qquad\qquad\qquad\qquad\quad \theta_k^i = j \\ \frac{P_D\lambda_B p\left(z_k^j|x_b\right)}{\lambda_{FA}+P_D\lambda_B p\left(z_k^j|x_b\right)}, \theta_k^{N_{k-1}+j} = N_{k-1}+j. \end{cases} \tag{31}$$

The posterior probability of target existence $p(\chi_k^i|Z_k)$ and the existence-conditioned marginal association probability $p(\theta_k^i|\chi_k^i, Z_k)$ can then be calculated using the Bayesian rule as

$$\begin{aligned} p\left(\chi_k^i\left|Z_k\right.\right) &= \sum_{\theta_k^i} p\left(\theta_k^i, \chi_k^i\left|Z_k\right.\right) \\ p\left(\theta_k^i\left|\chi_k^i, Z_k\right.\right) &= \frac{p\left(\theta_k^i, \chi_k^i\left|Z_k\right.\right)}{p\left(\chi_k^i\left|Z_k\right.\right)}. \end{aligned} \tag{32}$$

Using $p(\chi_k^i|Z_k)$ and $p(\theta_k^i|\chi_k^i, Z_k)$, we can now perform track update for the proposed JPDA using (9).

B. Efficient Computation of $p(\theta_k^i|Z_k)$

It is clear that the exact brute force solution through full enumeration (28) is computationally intractable except for a few simple cases because of the combinatorial nature. For efficient implementation, we utilize the stochastic Gibbs sampling to approximate the marginal association probability. The Gibbs sampling algorithm enables fast calculation of the marginal probability with ignorable performance sacrifice [22], [23]. We generate sufficient samples of Θ_k by Gibbs sampling and then it is straightforward to approximate the marginal probability by the event occurrence. Based on the theory of Gibbs sampling, the transition kernel from one joint event $\Theta_k = \left(\theta_k^1, \ldots, \theta_k^{N_{k-1}}\right)$ to another joint event $\bar{\Theta}_k = \left(\bar{\theta}_k^1, \ldots, \bar{\theta}_k^{N_{k-1}}\right)$ is given by

$$\pi\left(\bar{\Theta}_k\left|\Theta_k\right.\right) = \prod_{m=1}^{N_{k-1}} \pi_m\left(\bar{\theta}_k^m\left|\bar{\theta}_k^1, \ldots, \bar{\theta}_k^{m-1}, \theta_k^{m+1}, \ldots, \theta_k^{N_{k-1}}\right.\right) \tag{33}$$

where p_m is determined by

$$\begin{aligned} &\pi_m\left(\bar{\theta}_k^m\left|\bar{\theta}_k^1, \ldots, \bar{\theta}_k^{m-1}, \theta_k^{m+1}, \ldots, \theta_k^{N_{k-1}}\right.\right) \\ &= \frac{p\left(\bar{\theta}_k^1, \ldots, \bar{\theta}_k^m, \theta_k^{m+1}, \ldots, \theta_k^{N_{k-1}}\right)}{p\left(\bar{\theta}_k^1, \ldots, \bar{\theta}_k^{m-1}, \theta_k^{m+1}, \ldots, \theta_k^{N_{k-1}}\right)} \\ &\propto p\left(\bar{\theta}_k^1, \ldots, \bar{\theta}_k^m, \theta_k^{m+1}, \ldots, \theta_k^{N_{k-1}}\right) \\ &\propto p\left(\bar{\theta}_k^m\right) \end{aligned} \tag{34}$$

where the last line is obtained based the basic assumption that each single association hypothesis is independent of each other utilized in JPDA.

The preceding equation reveals that the Gibbs sampling only requires the individual prior association distribution, thus, avoiding full enumeration in the marginalization. The prior distribution of an individual association is given by

$$p\left(\theta_k^i\right) \propto \begin{cases} 1 - P_D p\left(\chi_k^i\left|Z_{k-1}\right.\right), & \theta_k^i = 0 \\ P_D p\left(\chi_k^i\left|Z_{k-1}\right.\right) p\left(z_k^j\left|x_{k|k-1}^i\right.\right), & \theta_k^i = j \\ \lambda_{FA} + P_D\lambda_B p\left(z_k^j\left|x_b\right.\right), & \theta_k^i = N_{k-1}+j. \end{cases} \tag{35}$$

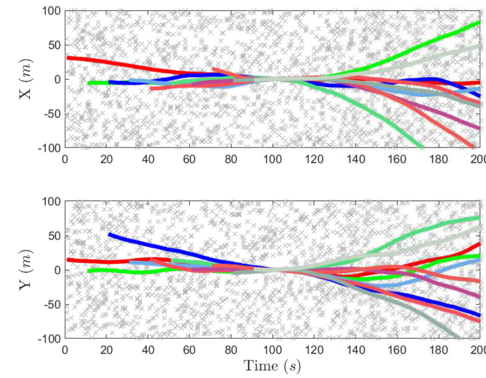


Fig. 2. Sample ground truth and measurements (grey “x: for measurements and color lines for targets).

Given the joint event Θ_k, a joint event $\bar{\Theta}_k$ can, therefore, be recursively sampled as

$$\bar{\theta}_k^m \sim \pi_m\left(\bar{\theta}_k^m\left|\bar{\theta}_k^1, \ldots, \bar{\theta}_k^{m-1}, \theta_k^{m+1}, \ldots, \theta_k^{N_{k-1}}\right.\right). \tag{36}$$

Then, the Gibbs samples will exponentially converge to the stationary distribution $p(\Theta_k|Z_k)$ as [22], [23]

$$\left|\pi^n\left(\bar{\Theta}_k\left|\Theta_k\right.\right) - p\left(\bar{\Theta}_k\left|Z_k\right.\right)\right| \le (1-2\beta)^{\lfloor n/2\rfloor} \tag{37}$$

where $\pi^n(\bar{\Theta}_k|\Theta_k)$ denotes the nth power of transition kernel $\pi(\bar{\Theta}_k|\Theta_k)$, $\beta = \min \pi^2(\bar{\Theta}_k|\Theta_k) \in (0, 0.5]$ the least likely two-step transition probability.

V. Simulation Studies

In this section, the effectiveness of the proposed MTT algorithm is demonstrated through extensive empirical numerical simulations. Our experiments explore a very challenging scenario, involving ten closely moving targets. These four targets appear at time step from $t = 0$~s to $t = 90$~s at every 10~s interval. The state vector contains planar position and velocity. We assume that all targets share the same motion model and use the well-known constant velocity (CV) model for target prediction. The CV model is defined as

$$x_k = F_{\mathrm{CV}} x_{k-1} + G w_{k-1} \tag{38}$$

with

$$F_{\mathrm{CV}} \triangleq \mathbb{I}_{2\times 2} \otimes \begin{bmatrix} 1 & T \\ 0 & 1 \end{bmatrix}, \quad G \triangleq \left[T^2/2, T, T^2/2, T\right]^T \tag{39}$$

where $\mathbb{I}_{2\times 2}$ denotes the 2×2 identity matrix, $T = 1$~s the sampling period, and $w_k \sim \mathcal{N}(\cdot; 0, \sigma_v^2)$ the Gaussian process noise with $\sigma_v = 0.1$~m/s^2. The linear position measurements, generated with $P_D = 0.9$, are modeled by

$$z_k = \left[x_{T,k}, \quad y_{T,k}\right]^T + v_k \tag{40}$$

where $(x_{T,k}, y_{T,k})$ denotes target position, and $v_k \sim \mathcal{N}(\cdot; 0, R_k)$ the Gaussian measurement noise with $R_k = \sigma_z^2 \mathrm{diag}([1, 1])$. Snapshots of one sample trajectory for each target and measurements are depicted in Fig. 2. The clutter is assumed to be uniformly distributed in the surveillance region with its number being Poisson with ten average returns at each scan. In order to cover the entire region of interest, the target birth model is $\lambda_B(x) = \lambda_B \mathcal{N}(x; x_b, P_b)$ with $\lambda_B = 0.05$, $x_b = [0, 0, 0, 0]^T$, $P_b = \mathrm{diag}([100^2, 1, 100^2, 1])$. A tentative track is confirmed if the existence probability satisfies $p(\chi_k^i|Z_k) \ge 0.8$ and a confirmed track is deleted immediately once $p(\chi_k^i|Z_k) \le 0.1$.

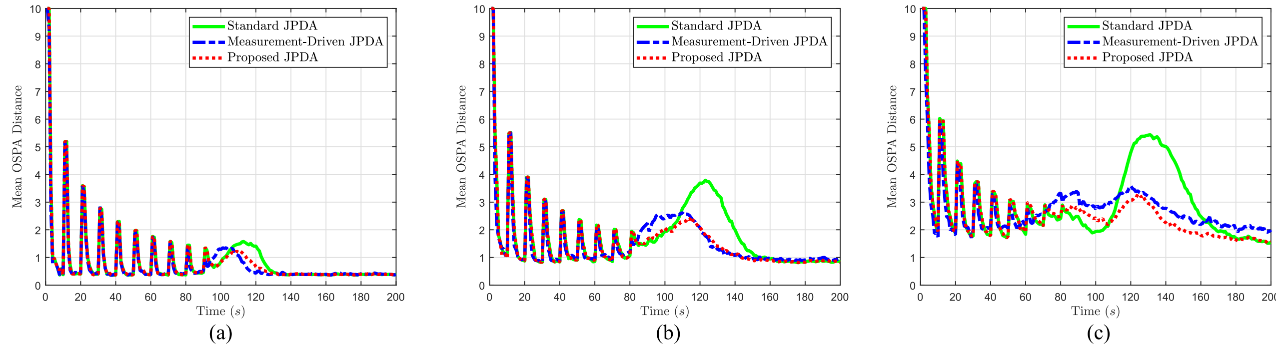


Fig. 3. Mean OSPA distance of 100 Monte Carlo simulations. (a) $\sigma_z = 0.3162$. (b) $\sigma_z = 1$. (c) $\sigma_z = 2.2361$.

We compare the proposed JPDA filter with standard JPDA and measurement-driven JPDA [13] for various cases. In order to make fair comparisons, the Gibbs sampling-aided marginalization is also utilized in other two algorithms. The optimal subpattern assignment (OSPA) distance metric [24] is considered here for overall evaluation of performance. The results of mean OSPA distance obtained by 100 Monte Carlo runs are shown in Fig. 3. The peaks of mean OSPA distance before 100˜s in this figure are resulted from track confirmation for target birth. Since all targets cross each other at around 100 s, all filters cannot correctly assign measurements to targets, which in turn results in performance degradation. Although the measurement-driven JPDA can also mitigate the multimodality issue in Gaussian mixture approximation, the resultant solution is sensitive to measurement noise, as confirmed by Fig. 3. It is clear from Fig. 3 that the proposed information-theoretic JPDA outperforms other two algorithms in terms of tracking accuracy and shows strong robustness against the variation of measurement noise in all cases. The reason is that minimizing the proposed KLD provides approximate mode selection in Gaussian mixture reduction for updating JPDA and, therefore, can alleviate the multimodality issue when targets are moving closely around 100 s. Therefore, the proposed information-theoretic approach is a competitive alternative to the original JPDA and can be expected to address the Gaussian mixture approximation problem in related methods.

VI. Conclusion

We have proposed a novel information-theoretic JPDA algorithm that can be utilized to track varying number of targets in this article. The proposed JPDA filter is derived through the minimization of a weighted reverse KLD to approximate the posterior Gaussian mixture PDF. Analytical performance bounds are also derived to support the proposed approach. Extensive empirical simulations clearly validate the effectiveness of the proposed approach.

Appendix A
Proof of Proposition 1

According to the log-sum inequality, we have

$$\mathcal{D}_{\mathrm{KL}}\left(p^i \left\| \sum_{j=0}^{M_k} \beta_j^i p_j^i \right.\right) = \int p^i\left(x^i\right) \ln \frac{p^i\left(x^i\right)}{\sum_{j=0}^{M_k} \beta_j^i p_j^i\left(x^i\right)} dx^i$$

$$\leq \int \sum_{j=0}^{M_k} a_j p^i\left(x^i\right) \ln \frac{a_j p^i\left(x^i\right)}{\beta_j^i p_j^i\left(x^i\right)} dx^i$$

$$= \sum_{j=0}^{M_k} a_j \int p^i\left(x^i\right) \ln \frac{p^i\left(x^i\right)}{p_j^i\left(x^i\right)} dx^i + \sum_{j=0}^{M_k} a_j \ln \frac{a_j}{\beta_j^i}$$

$$= \sum_{j=0}^{M_k} a_j \mathcal{D}_{\mathrm{KL}}\left(p^i \left\| p_j^i\right.\right) + \sum_{j=0}^{M_k} a_j \ln \frac{a_j}{\beta_j^i} \tag{41}$$

where $0 \leq a_j \leq 1$ and $\sum_{j=0}^{M_k} a_j = 1$, and the equality holds if and only if

$$\frac{a_{j'} p^i\left(x^i\right)}{\beta_{j'}^i p_{j'}^i\left(x^i\right)} = \frac{a_j p^i\left(x^i\right)}{\beta_j^i p_j^i\left(x^i\right)} \quad \forall j' \neq j. \tag{42}$$

Then, Proposition 1 can be proved by choosing $a_j = \beta_j^i$.

Appendix B
Proof of Proposition 2

From (7), one can imply that

$$\sum_{j=0}^{M_k} \beta_j^i \mathcal{D}_{\mathrm{KL}}\left(p^i \left\| p_j^i\right.\right) = \sum_{j=0}^{M_k} \beta_j^i \int p^i\left(x^i\right) \ln \frac{p^i\left(x^i\right)}{p_j^i\left(x^i\right)} dx^i$$

$$= \int p^i\left(x^i\right) \ln \frac{p^i\left(x^i\right)}{\prod_{j=0}^{M_k} \left[p_j^i\left(x^i\right)\right]^{\beta_j^i}} dx^i$$

$$= \int p^i\left(x^i\right) \ln \frac{p^i\left(x^i\right)}{c p_*^i\left(x^i\right)} dx^i$$

$$= \mathcal{D}_{\mathrm{KL}}\left(p^i \left\| p_*^i\right.\right) - \ln c \tag{43}$$

where $p_*^i = \frac{1}{c} \prod_{j=0}^{M_k} [p_j^i(x^i)]^{\beta_j^i}$ and c is a constant.

Since $\mathcal{D}_{\mathrm{KL}}(p^i \| p_*^i) \geq 0$ and the equality holds if and only if $p^i = p_*^i$, the solution of optimization problem (7) is $p^i = p_*^i$. Assume that $p_j^i(x^i)$ takes Gaussian distribution $\mathcal{N}(x^i; x_{k|k}^{i,j}, P_{k|k}^{i,j})$, then, we have

$$\left[p_j^i\left(x^i\right)\right]^{\beta_j^i} \propto \mathcal{N}\left(x^i; x_{k|k}^{i,j}, P_{k|k}^{i,j} / \beta_j^i\right)$$

$$\prod_{j=0}^{M_k} p_j^i\left(x^i\right) \propto \mathcal{N}\left(x^i; \mu^i, \Sigma^i\right) \tag{44}$$

where $\mu^i = \Sigma^i \sum_{j=0}^{M_k} (P_{k|k}^{i,j})^{-1} x_{k|k}^{i,j}$, $\Sigma^i = (\sum_{j=0}^{M_k} (P_{k|k}^{i,j})^{-1})^{-1}$.

Substituting (44) into p_*^i yields $p_*^i \propto \mathcal{N}(x^i; x_{k|k}^i, P_{k|k}^i)$. Since $\int p_*^i(x^i) dx^i = 1$ and $\int \mathcal{N}(x^i; x_{k|k}^i, P_{k|k}^i) dx^i = 1$, it is easy to verify that $p_*^i = \mathcal{N}(x^i; x_{k|k}^i, P_{k|k}^i)$.

References

[1] D. L. Alspach, "A Gaussian sum approach to the multi-target identification-tracking problem," *Automatica*, vol. 11, no. 3, pp. 285–296, 1975.

[2] A. Roy and D. Mitra, "Unscented-Kalman-filter-based multitarget tracking algorithms for airborne surveillance application," *J. Guid., Control, Dyn.*, vol. 39, no. 9, pp. 1949–1966, 2016.

[3] A. Pak, J. Correa, and M. Adams, "Robust joint target detection and tracking for space situational awareness," *J. Guid., Control, Dyn.*, vol. 41, no. 1, pp. 119–136, 2018.

[4] C. S. Lee, D. E. Clark, and J. Salvi, "Slam with dynamic targets via single-cluster PhD filtering," *IEEE J. Sel. Topics Signal Process.*, vol. 7, no. 3, pp. 543–552, Jun. 2013.

[5] C. Kim, F. Li, A. Ciptadi, and J. M. Rehg, "Multiple hypothesis tracking revisited," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2015, pp. 4696–4704.

[6] D. Reid, "An algorithm for tracking multiple targets," *IEEE Trans. Autom. Control*, vol. 24, no. 6, pp. 843–854, Dec. 1979.

[7] I. J. Cox and S. L. Hingorani, "An efficient implementation of Reid's multiple hypothesis tracking algorithm and its evaluation for the purpose of visual tracking," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 18, no. 2, pp. 138–150, Feb. 1996.

[8] T. Fortmann, Y. Bar-Shalom, and M. Scheffe, "Sonar tracking of multiple targets using joint probabilistic data association," *IEEE J. Ocean. Eng.*, vol. OE-8, no. 3, pp. 173–184, Jul. 1983.

[9] S. Oh, S. Russell, and S. Sastry, "Markov chain Monte Carlo data association for general multiple-target tracking problems," in *Proc. 43rd IEEE Conf. Decis. Control*, 2004, pp. 735–742.

[10] R. J. Fitzgerald, "Development of practical PDA logic for multitarget tracking by microprocessor," in *Proc. Amer. Control Conf.*, 1986, pp. 889–898.

[11] L. Svensson, D. Svensson, M. Guerriero, and P. Willett, "Set JPDA filter for multitarget tracking," *IEEE Trans. Signal Process.*, vol. 59, no. 10, pp. 4677–4691, Oct. 2011.

[12] R. A. Lau and J. L. Williams, "A structured mean field approach for existence-based multiple target tracking," in *Proc. 19th Int. Conf. Inf. Fusion*, 2016, pp. 1111–1118.

[13] J. L. Williams, "Marginal multi-Bernoulli filters: Rfs derivation of MHT, JIPDA, and association-based member," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 51, no. 3, pp. 1664–1687, Jul. 2015.

[14] C. M. Bishop, *Pattern Recognition and Machine Learning*. Berlin, Germany: Springer, 2006.

[15] S. Challa, M. R. Morelande, D. Mušicki, and R. J. Evans, *Fundamentals of Object Tracking*. Cambridge, MA, USA: Cambridge Univ. Press, 2011.

[16] Y. Bar-Shalom, X. R. Li, and T. Kirubarajan, *Estimation With Applications to Tracking and Navigation: Theory Algorithms and Software*. Hoboken, NJ, USA: Wiley, 2004.

[17] K. L. Mengersen, C. Robert, and M. Titterington, *Mixtures: Estimation and Applications*, vol. 896. Hoboken, NJ, USA: Wiley, 2011.

[18] A. R. Runnalls, "Kullback-Leibler approach to Gaussian mixture reduction," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 43, no. 3, pp. 989–999, Jul. 2007.

[19] S. He, H.-S. Shin, S. Xu, and A. Tsourdos, "Distributed estimation over a low-cost sensor network: A review of state-of-the-art," *Inf. Fusion*, vol. 54, pp. 21–43, 2020.

[20] W. Niehsen, "Information fusion based on fast covariance intersection filtering," in *Proc. 5th Int. Conf. Inf. Fusion*, 2002, pp. 901–904.

[21] J. Deyst and C. Price, "Conditions for asymptotic stability of the discrete minimum-variance linear estimator," *IEEE Trans. Autom. Control*, vol. 13, no. 6, pp. 702–705, Dec. 1968.

[22] B.-N. Vo, B.-T. Vo, and H. G. Hoang, "An efficient implementation of the generalized labeled multi-Bernoulli filter," *IEEE Trans. Signal Process.*, vol. 65, no. 8, pp. 1975–1987, Apr. 2017.

[23] S. He, H.-S. Shin, and A. Tsourdos, "Track-oriented multiple hypothesis tracking based on tabu search and Gibbs sampling," *IEEE Sensors J.*, vol. 18, no. 1, pp. 328–339, Jan. 2017.

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