This article has been accepted for publication in IEEE Transactions on Automatic Control. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TAC.2024.3375254

A Distributed Bayesian Data Fusion Algorithm with Uniform Consistency

Yingke Li, Enlu Zhou and Fumin Zhang

***Abstract*— Distributed Data Fusion (DDF) methods which possess guaranteed performance for ad-hoc and arbitrarily connected networks empower more scalable, flexible and robust information fusion for multi-robot sensor networks. This paper proposes a novel distributed Bayesian data fusion algorithm, which ensures uniform consistency, i.e., all the locally estimated distributions converge to the true distribution, for arbitrary periodically connected communication graphs. Conservative fusion via the Weighted Exponential Product (WEP) rule is utilized to combat inconsistencies that arise from double-counting common information between fusion agents, and the WEP fusion weight is chosen based on the dynamic communication network topology. The uniform consistency of the proposed algorithm is rigorously proved, and the cooperative consistency conditions that guarantee uniform consistency have been explicitly identified. The performance and convergence properties of the proposed algorithm are validated through simulations.**

***Index Terms*— Bayesian Learning, Bayesian Consistency, Distributed Data Fusion**

I. Introduction

Fusing information from an ensemble of noisy data streams in a scalable, flexible and robust way is critical to many multi-robot systems operating in uncertain dynamic environments, especially for applications such as collaborative mapping for exploration, target search/tracking for surveillance, and futuristic unmanned urban transport systems [1]–[7]. In those systems, the communication between networked robots can be highly variable due to time-varying environmental factors: sensor nodes could fail, or be added, or perhaps the network could be reconfigured as sensors move around and communication proximity changes. Hence, the feasibility of any centralized data fusion method reliant on full communication graph connectivity is compromised, rendering it fragile over extended real-world operations. As a sequel, effective distributed data fusion methods should possess guaranteed performance for ad-hoc and arbitrarily connected networks, which provide scalability for fusion agents to join and drop off the network, flexibility to allow agents to join at any point, and robustness to ensure connectivity of the network even when there are multiple failures of links or agents [8]–[17].

In general network topology, specifically circular topology, there may exist unknown redundant information that is propagated through the network. For example, in a circular topology, information can be passed in both directions around the circle, but as it comes back to the original node, that node cannot infer the difference because of the underlying assumption of unique information between nodes. This concept is called *rumor propagation* where nodes have *common information* [1]. The double-counting of common information between fusion agents may harm the statistical consistency during the data fusion process in the network. Consider a worst-case scenario, if an outlier measurement is propagated through a circular network and accounted for infinitely number of times, all the estimators in the network will be polluted. To avoid such inconsistencies due to rumor propagation in general network topology, conservative data fusion techniques are developed via the *Weighted Exponential Product* (WEP) rule, which can maintain consistent estimates for arbitrary probability distributions with unknown correlation. [18]–[25].

The most critical component in WEP rule based distributed data fusion approach is the choice of the fusion weights, which determines the performance of the data fusion process. Many different information-theoretic metrics have been proposed to determine the optimal WEP fusion weight, for example, *Chernoff Fusion*, *Shannon Fusion*, *Bhattacharyya Fusion* [26], *Generalized Information Weighted Chernoff Fusion*, and *Minimum Information Loss Weight Fusion* [18]. However, those information-based methods assume the homogeneity of different data streams in the distributed sensor network, and thus lack the flexibility to incorporate any unique properties of individual sensor models or communication network topology. Furthermore, none of those approaches can provide any guarantees pertaining to the consistency of the data fusion process, i.e., whether the fused estimation can effectively distill the ground truth from a complex mixture of noisy information streams.

In this paper, we proposed a novel strategy for choosing the WEP fusion weights that offers the flexibility to explicitly take various facets of sensor network properties into consideration. In this approach, the information exchange protocol between an agent and all of its neighbors on the network is encoded by the trust that one sensor places on the estimates of the other, and the WEP fusion weights are determined based on the dynamic trust between sensors. The introduction of a dynamic trust network, whose adjacent matrix represents the mutual trust between nodes, provides the opportunity to integrate distinct attributes of individual sensor models or communication network topologies by deploying different strategies when determining trust. For example, in a heterogeneous sensor network, the trust can depend on the relative reliability of different types of sensors. Sensors that have more accurate or frequent measurements should contribute more to the whole network. It is also possible to reject faulty sensors by diminishing the trust in the sensors that show dramatic disparity with other sensors.

Taking advantage of the well-established properties in matrix theory, algebraic graph theory, and control theory [13], [27], the proposed algorithm also demonstrates its merits by providing the assurance of uniform consistency, i.e., all the locally estimated distributions converge to the true distribution, for arbitrary periodically connected communication graphs. To the best of our knowledge, this guarantee hasn't appeared in the existing literature on distributed data fusion methods. In this paper, we explicitly identify the cooperative consistency conditions which guarantee uniform consistency. Compared to the consistency conditions for a single agent, which was stated in our previous work [28], they are weaker conditions since it is easier for the unknown parameter to satisfy the distinguishable requirement within the joint sampled dataset for all the agents than for a single agent. The cooperative consistency conditions greatly reduce the burden of computation and mobility for each agent, since the requirement of the richness in sampled data is now distributed to multiple agents.

Yingke Li and Fumin Zhang are with school of Electrical and Computer Engineering, Georgia Institute of Technology, Atlanta, GA, USA yli3225,fumin@gatech.edu

Enlu Zhou is with School of Industrial and Systems Engineering, Georgia Institute of Technology, Atlanta, GA, USA ezhou30@gatech.edu

The research work is supported by ONR grants N00014-19-1-2556 and N00014-19-1-2266; AFOSR grants FA9550-19-1-0283 and FA9550-22-1-0244; NSF grants GCR-1934836, CNS-2016582, ITE-2137798 and DMS-2053489; and NOAA grant NA16NOS0120028.

II. Background on Bayesian Data Fusion

Consider a network of N heterogeneous sensors simultaneously collecting information to estimate an unknown parameter $\theta^* \in \Theta$ in a workspace $\mathcal{X}$. For example, a group of robots search for a target whose location θ^* is unknown to the robots. Assume that the parameter space Θ and workspace $\mathcal{X}$ are all finite and discrete. A Bayesian estimation of θ^* can be represented by a random variable $\theta \in \Theta$ which obeys a probability distribution $\pi : \mathcal{X} \to [0,1]$ such that $\pi(\theta)$ is the probability that $\theta = \theta^*$.

A. Measurement Model

Let x_k^i denote the location of the ith sensor at the kth time instant, and z_k^i is the measurement taken by the ith sensor at its current location x_k^i. The collected measurement z_k^i of each sensor is modeled as a noisy observation of the unknown parameter θ, which depends on the sensor's type and its current location x_k^i. Then the probability distribution of z_k^i can be described by the measurement model $p^i(z_k^i|\theta, x_k^i)$.

We assume that the sensors in the network are memoryless and independent, which implies that the measurements collected by different sensors at different time instants are conditionally independent given the current location and time.

Assumption II.1 (Memoryless and independent sensors)**.** *Let $z_{1:k}^i$ be the measurements collected by the ith sensor at locations $x_{1:k}^i$. Then $p(z_k^{1:N}|\theta, z_{1:k-1}^{1:N}, x_{1:k}^{1:N}) = \prod_{i=1}^N p^i(z_k^i|\theta, z_{1:k-1}^{1:N}, x_{1:k}^{1:N}) = \prod_{i=1}^N p^i(z_k^i|\theta, x_k^i)$.*

The likelihood function $p^i(z_k^i|\theta, x_k^i)$ can be obtained from the ith sensor's known measurement model.

B. Centralized Bayesian Data Fusion

The centralized multi-sensor Bayesian data fusion algorithm updates the probability distribution of the unknown parameter using the measurements collected from all the sensors.

Define $\pi_k(\theta) = p(\theta|z_{1:k}^{1:N}, x_{1:k}^{1:N})$ as the centralized posterior distribution of the unknown parameter after k time instants. According to the *Bayes theorem*:

$$\begin{aligned}\pi_k(\theta) &= p(\theta|z_{1:k}^{1:N}, x_{1:k}^{1:N})\\ &= \frac{p(\theta|z_{1:k-1}^{1:N}, x_{1:k}^{1:N})p(z_k^{1:N}|\theta, z_{1:k-1}^{1:N}, x_{1:k}^{1:N})}{\sum_\theta p(\theta|z_{1:k-1}^{1:N}, x_{1:k}^{1:N})p(z_k^{1:N}|\theta, z_{1:k-1}^{1:N}, x_{1:k}^{1:N})}\\ &= \frac{p(\theta|z_{1:k-1}^{1:N}, x_{1:k}^{1:N})\prod_{i=1}^N p^i(z_k^i|\theta, x_k^i)}{\sum_\theta p(\theta|z_{1:k-1}^{1:N}, x_{1:k}^{1:N})\prod_{i=1}^N p^i(z_k^i|\theta, x_k^i)},\end{aligned}$$

where the last equality holds according to the assumption of memoryless and independent sensors (Assumption II.1).

Since sensors moving to new locations does not affect their information on the unknown parameter until new measurements are taken, we have

$$p(\theta|z_{1:k-1}^{1:N}, x_{1:k}^{1:N}) = p(\theta|z_{1:k-1}^{1:N}, x_{1:k-1}^{1:N}) = \pi_{k-1}(\theta).$$

Therefore, the recursive Bayesian update rule simplifies as

$$\pi_k(\theta) = \frac{\pi_{k-1}(\theta)\prod_{i=1}^N p^i(z_k^i|\theta, x_k^i)}{\sum_\theta \pi_{k-1}(\theta)\prod_{i=1}^N p^i(z_k^i|\theta, x_k^i)}, \quad (1)$$

which means that the computation of the posterior distribution of the current iteration $\pi_k(\theta)$ requires the posterior distribution of the previous iteration $\pi_{k-1}(\theta)$ and all the measurements collected at the current iteration k.

The centralized multi-sensor Bayesian data fusion algorithm is considered *Bayesian-optimal* since the centralized posterior distribution integrates all the available information expressed by probabilities, which can be utilized as a benchmark to make performance comparisons with distributed Bayesian data fusion algorithms.

C. Distributed Bayesian Data Fusion

If all the sensors in the network are perfectly connected on a complete communication graph (i.e., each agent could communicate instantaneously with every other agent without any loss of information in the communication links), then all the sensors can exchange their local likelihood functions and deploy the centralized multi-sensor Bayesian data fusion algorithm to estimate the Bayesian-optimal posterior probability distribution of the unknown parameter. However, communication between networked mobile sensors can be highly variable and brittle due to dynamic and unstructured environmental factors. Therefore, for ad-hoc and arbitrary network topologies, we resort to distributed Bayesian data fusion approaches where information exchanges only happen among neighbors.

The WEP rule [12], also known as the *normalized weighted geometric mean* or *logarithmic opinion pool*, is a popular conservative data fusion technique that is ideal for fusing arbitrarily distributed estimates in ad-hoc communication network topologies. It is considered conservative since it is able to efficiently prevent the double-counting of common information. To illustrate this point in a clear way, consider two information sources Y^1 and Y^2 with associated conditional PDFs $p^i(\theta) = p(\theta|Y^i)$ for $i = 1, 2$. In general, the collective information set $Y^1 \cup Y^2$ can be decomposed into the union of three disjoint (independent) information sets as follows $Y^1 \cup Y^2 = (Y^1 \cap Y^2) \cup (Y^1 \backslash Y^2) \cup (Y^2 \backslash Y^1)$. Hence, the optimal fusion of $p^1(\theta)$ and $p^2(\theta)$ is given as

$$\begin{aligned}p^{12}(\theta) &\propto p(\theta|Y^1 \cup Y^2)\\ &\propto p(\theta|Y^1 \cap Y^2)p(\theta|Y^1\backslash Y^2)p(\theta|Y^2\backslash Y^1),\end{aligned}$$

if the conditional PDF $p(\theta|Y^1 \cap Y^2)$ were known.

However, in general network topology, since the nodes usually do not have full knowledge of the network topology, it is difficult to discriminate the common information $p(\theta|Y^1 \cap Y^2)$. If we combine the two conditional PDFs naively as

$$\begin{aligned}\tilde{p}^{12}(\theta) &\propto p(\theta|Y^1)p(\theta|Y^2)\\ &\propto p^2(\theta|Y^1 \cap Y^2)p(\theta|Y^1\backslash Y^2)p(\theta|Y^2\backslash Y^1),\end{aligned}$$

this combination leads to a double-counting of $p(\theta|Y^1 \cap Y^2)$.

Instead, the WEP fusion rule can provide a principled way to combine two general distributions in a statistically consistent way despite the fact that their mutual dependency is unknown:

$$\begin{aligned}\bar{p}^{12}(\theta) &\propto [p^1(\theta)]^\omega[p^2(\theta)]^{1-\omega}\\ &\propto p(\theta|Y^1 \cap Y^2)[p(\theta|Y^1\backslash Y^2)]^\omega[p(\theta|Y^2\backslash Y^1)]^{1-\omega},\end{aligned}$$

where $\omega \in [0,1]$ is the fusion parameter. Note that the common information $p(\theta|Y^1 \cap Y^2)$ was handled correctly for any choice of $\omega \in [0,1]$, implying that the nodes can be connected in any network topology, and $p(\theta|Y^1 \cap Y^2)$ does not have to be tracked locally.

However, this conservative fusion also loses at least some new exclusive information from $p(\theta|Y^1\backslash Y^2)$ and $p(\theta|Y^2\backslash Y^1)$ during the process. Thus, while WEP fusion has the advantage of working for arbitrary networks and distributions, this flexibility comes at the expense of conservative information loss. As a result, the selection of the WEP fusion parameter ω is of great importance to balance this tradeoff. For example, different choice of the WEP fusion parameter decides how much partial of the exclusive information from each local estimate will be preserved.

Let $\pi_k^i(\theta)$ be the locally estimated posterior distribution of sensor i at time instant k. Taking advantage of the WEP rule, the distributed Bayesian data fusion approach updates the local estimation of each sensor based on its own measurement $p^i(z_k^i|\theta, x_k^i)$, as well as all the local estimations from its neighbors:

$$\pi_k^i(\theta) = \frac{p^i(z_k^i|\theta, x_k^i) \prod_{j=1}^N [\pi_{k-1}^j(\theta)]^{\alpha_k^{i,j}}}{\sum_\theta p^i(z_k^i|\theta, x_k^i) \prod_{j=1}^N [\pi_{k-1}^j(\theta)]^{\alpha_k^{i,j}}}; \tag{2}$$

where the fusion weights satisfy $\sum_{j=1}^N \alpha_k^{i,j} = 1, \alpha_k^{i,j} \geq 0, \forall j \in \{1, 2, \dots, N\}$.

III. Problem Formulation

The most critical component in WEP rule based distributed data fusion approach is the choice of the fusion weights $\alpha_k^{i,j}$. Most existing approaches determine the WEP fusion weights by optimizing various information-theoretic metrics, for example, *Chernoff Fusion*, *Shannon Fusion*, *Bhattacharyya Fusion*, *Generalized Information Weighted Chernoff Fusion*, and *Minimum Information Loss Weight Fusion*. However, those methods lack the flexibility to integrate any distinct attributes of individual sensor models or communication network topologies, nor can they provide any convergence guarantees of the data fusion process.

Therefore, the objective of this paper is to devise a novel strategy for choosing the WEP fusion weights that (i) offers the flexibility to explicitly take various facets of sensor network properties into consideration, and also (ii) provides theoretically assured performance guarantees.

Dynamic Trust Network: The time-varying communication network topology of the sensor network is denoted by the directed graph $G_k = (V, E_k)$ with the set of nodes $V = \{1, 2, \dots, N\}$ and edges $E_k \subseteq V \times V$. The edge $(i, j) \in E_k$ if and only if the ith sensor receives information from the jth sensor at the kth time instant. The neighbors of the ith sensor are denoted by $\mathcal{N}_k^i = \{j \in V : (i, j) \in E_k\}$. The trust that one sensor places on the estimates of the other is encoded by the adjacency matrix $A_k \in \mathbb{R}^{N\times N}$ of the communication network topology G_k, which can also be reckoned as the *trust matrix*. In this paper, we utilize $M[i, j]$ to infer the element at the ith row and jth column of a matrix M. Then $A_k[i, j] \neq 0$ if and only if $j \in \mathcal{N}_k^i$.

Definition III.1 (Periodically Connected Graph)**.** *The dynamic graph is* ***periodically connected*** *with a period $T \geq 1$ if the union of all graphs over a sequence of intervals $[k, k+T)$ is a connected graph, i.e., $G_{k:k+T} = \bigcup_{t=0}^{T-1} G_{k+t}$ is connected for all time instants $k \in \mathbb{N}$.*

Uniform Consistency: In statistics, a consistent estimator is one for which, when the estimate is considered as a random variable indexed by the number n of items in the data set, as n increases the estimates converge in probability to the value that the estimator is designed to estimate. For multiple estimators that formulate a network, we define the *uniform consistency* of networked estimators as:

Definition III.2 (Uniform Consistency)**.** *The Bayesian estimators in a sensor network are* ***uniformly consistent*** *if each agent's locally estimated distribution of θ, $\pi_k^i(\theta), \forall i \in \{1, 2, \dots, N\}$, all converge to the true distribution $\pi^*(\theta)$, with probability 1 (w.p.1).*

Problem Statement:

Problem 1 (Uniformly Consistent Distributed Bayesian Data Fusion)**.** *Given any periodically connected sensor network, find a distributed strategy to dynamically decide the fusion weight $\alpha_k^{i,j}$ between any neighboring sensors based on the time-varying communication network topology G_k, which ensures uniform consistency of the networked estimators.*

IV. Uniformly Consistent Distributed Bayesian Data Fusion

A. Algebraic Graph Theory

In order to derive and analyze the distributed Bayesian data fusion algorithm, we first introduce some important concepts and results in the field of *algebraic graph theory* [27].

Definition IV.1 (Laplacian Matrix)**.** *The* ***Laplacian matrix*** *of a graph G is defined as $L = D - A$, where A is the adjacency matrix of G, and $D = diag(d_1, d_2, \dots, d_N)$ is the degree matrix of G with elements $d_i = \sum_{j\neq i} A[i, j]$ and zero off-diagonal elements.*

Definition IV.2 (Perron Matrix)**.** *The* ***Perron matrix*** *of a graph G with parameter ϵ is defined as $P = I - \epsilon L$, where I is the identity matrix, L is the* ***Laplacian matrix*** *of the graph G, and $\epsilon \geq 0$ is the discrete step size.*

Therefore, the Perron matrix can be written as $P = (I - \epsilon D) + \epsilon A$, where $(I - \epsilon D)$ is a diagonal matrix. Each non-zero element in the diagonal matrix $1 - \epsilon d_i$ encodes the weight that each node i puts on its self-information, and $\epsilon A[i, j]$ represents its reliance on the information from another node j. It possesses numerous favorable properties in the field of algebraic graph theory.

Three important types of *non-negative matrices* are irreducible, stochastic, and primitive (or ergodic) matrices. A matrix is *irreducible* if its associated graph is strongly connected. A non-negative matrix is called row (or column) *stochastic* if all of its row-sums (or column-sums) are 1. An irreducible stochastic matrix is *primitive* if it has only one eigenvalue with maximum modulus.

Lemma IV.1. *Let G be a directed graph with N nodes and maximum degree $\Delta = \max_i(\sum_{j\neq i} A[i, j])$. Then the Perron matrix P with parameter $\epsilon \in (0, 1/\Delta]$ satisfies the following properties:*

(i) *P is a row stochastic non-negative matrix with a trivial eigenvalue of 1;*
(ii) *All eigenvalues of P are in a unit circle;*
(iii) *If G is strongly connected and $0 < \epsilon < 1/\Delta$, then P is a primitive matrix.*

Proof. See the proof of *Lemma 3* in [27]. □

Moreover, for periodically connected graphs, there exists a similar property to (iii):

Lemma IV.2. *Let $\{t_1, t_2, \dots, t_m\}$ be a set of indices for which the union of $G_{t_1}, G_{t_2}, \dots, G_{t_m}$ is a connected graph. Then the Perron matrix product $P_{t_1} P_{t_2} \cdots P_{t_m}$ is primitive.*

Proof. See the proof of *Lemma 1* in [15]. □

B. Uniformly Consistent Distributed Bayesian Data Fusion

Taking advantage of the well-established properties of *the WEP fusion rule* and *algebraic graph theory*, we design a *uniformly consistent distributed Bayesian data fusion algorithm*, as demonstrated in Algorithm 1.

For each sensor at each time step, the update of the locally estimated probability distribution is composed of two steps:

(i) **Fusion Step**. Each sensor communicates its current locally estimated probability distribution to its neighbors, and fuses all the received distributions based on the WEP rule, where the fusion parameters are chosen according to the Perron matrix P_k of the current communication graph G_k:

$$\hat{\pi}_k^i(\theta) = \frac{\prod_{j\in\{i\}\cup\mathcal{N}_k^i} [\pi_{k-1}^j(\theta)]^{P_k[i,j]}}{\sum_\theta \prod_{j\in\{i\}\cup\mathcal{N}_k^i} [\pi_{k-1}^j(\theta)]^{P_k[i,j]}}; \tag{3}$$

(ii) **Update Step**. Each sensor takes a local measurement and updates the local posterior distribution according to the likelihood function of the new measurement:

$$\pi_k^i(\theta) = \frac{p^i(z_k^i|\theta, x_k^i)\hat{\pi}_k^i(\theta)}{\sum_\theta p^i(z_k^i|\theta, x_k^i)\hat{\pi}_k^i(\theta)}; \tag{4}$$

Combining the two steps together, the update rule of each locally estimated probability distribution can be written as:

$$\pi_k^i(\theta) = \frac{p^i(z_k^i|\theta, x_k^i)\prod_{j=1}^N [\pi_{k-1}^j(\theta)]^{P_k[i,j]}}{\sum_\theta p^i(z_k^i|\theta, x_k^i)\prod_{j=1}^N [\pi_{k-1}^j(\theta)]^{P_k[i,j]}}, \tag{5}$$

since the Perron matrix P_k automatically selects the neighboring nodes that communicate with node i, i.e. $\forall i \neq j$, $P_k[i,j] \neq 0$ if and only if $A_k[i,j] \neq 0$.

Algorithm 1: Uniformly Consistent Distributed Bayesian Data Fusion Algorithm

For each sensor $i \in \{1, 2, \ldots, N\}$:
Initialize prior $\pi_0^i(\theta)$ with uniform probability distribution over the discretized parameter space Θ;
Initialize sensor position x_1^i;
Initialize time instant $k = 1$;
while *True* **do**
 Receive information from its neighbors and fuse to its locally estimated probability distribution:

$$\hat{\pi}_k^i(\theta) = \frac{\prod_{j\in\{i\}\cup\mathcal{N}_k^i}[\pi_{k-1}^j(\theta)]^{P_k[i,j]}}{\sum_\theta \prod_{j\in\{i\}\cup\mathcal{N}_k^i}[\pi_{k-1}^j(\theta)]^{P_k[i,j]}};$$

 Take a measurement z_k^i at the current location x_k^i;
 Update the local posterior distribution $\pi_k^i(\theta)$ as:

$$\pi_k^i(\theta) = \frac{p^i(z_k^i|\theta, x_k^i)\hat{\pi}_k^i(\theta)}{\sum_\theta p^i(z_k^i|\theta, x_k^i)\hat{\pi}_k^i(\theta)};$$

 Move to the next location x_{k+1};
 $k := k + 1$;
end

Remark 1. Note that no other side information, other than the current locally estimated probability distribution, is needed to communicate through the network. This means that sensors with local dynamics/states, even different types of sensors/dynamics/states, can effectively cooperate in the network. No knowledge of the local sensors/dynamics/states information is required to be sent across the network, which makes the network more scalable.

Remark 2. Moreover, a key advantage of this algorithm is that it can effectively deal with lost communication. For instance, sensors with communication limitations may lose connection due to disturbances in the environment and restore connection after a long period of time. If a sensor makes several measurements and then comes into contact with another sensor that has been out of contact for a long time, sharing the entire history would be costly in terms of both communication and computation required to fuse the measurements into each sensor's current estimation. Instead, with the proposed algorithm, only the current estimation needs to be maintained and shared across the network, since all the historical measurement information has already been fused into the current estimation.

Remark 3. It is also worth pointing out that the proposed algorithm does not require all the sensors to have the same frequency of measurements. If a sensor does not take a measurement at a certain time instant, then the likelihood function can be set as the uniform distribution, i.e., $p^i(z_k^i|\theta, x_k^i) = 1$. Then this agent's likelihood function does not change the locally estimated probability distribution because of the geometric nature of the fusion rule. Thus, the update of the estimation is only induced by receiving new information from all its neighbors.

V. Convergence Analysis

In this section, we analyze the convergence properties of the proposed algorithm, whose proof relies on *Wolfowitz's Lemma*:

Lemma V.1 (Wolfowitz's Lemma [29])**.** *Let $\mathcal{P} = \{P_1, P_2, \ldots, P_m\}$ be a finite set of primitive stochastic matrices such that for any sequence of matrices $P_{s_1}, P_{s_2}, \ldots, P_{s_k} \in \mathcal{P}$ with $k \geq 1$, the product $P_{s_k}\cdots P_{s_2}P_{s_1}$ is a primitive matrix. Then, for each infinite sequence of matrices, there exists a row vector α such that*

$$\lim_{k\to\infty} P_{s_k}\cdots P_{s_2}P_{s_1} = \mathbf{1}\alpha.$$

To facilitate the consistency proof, we assume that the true unknown parameter is a deterministic value, rather than a stochastic parameter sampled from a certain probability distribution:

Assumption V.1. *The true probability distribution of the unknown parameter is $\delta_{\theta^*}(\theta)$ where $\delta_{\theta^*}(\theta^*) = 1$ and $\delta_{\theta^*}(\theta) = 0$ for $\theta \neq \theta^*$.*

Remark 4. Note that this assumption is introduced only for the purpose of proving the consistency result since the estimated distribution can pinpoint the true parameter only if it is deterministic. For example, in the context of source seeking, we can determine the location of the source only if it is stationary. If the source has certain random movements, the only thing that can be done is to infer the probability distribution of the source rather than precisely identifying its location.

We now state the *cooperative consistency conditions*, which ensure uniform consistency for the proposed algorithm.

Theorem V.2 (Cooperative Consistency Conditions)**.** *The Bayesian estimators in a sensor network are* ***uniformly consistent*** *if the following* ***cooperative consistency conditions*** *are satisfied:*

(i) *The time-varying communication network topology of the sensor network is periodically connected;*
(ii) *The local prior distributions $\pi_0^i(\theta), i \in \{1, 2, \ldots, N\}$ have non-zero probability at θ^*;*
(iii) *Let $\Theta = \{\theta_1, \theta_2, \ldots, \theta_s\}$. If*

$$c_1 p^i(z|\theta_1, x) + c_2 p^i(z|\theta_2, x) + \cdots + c_s p^i(z|\theta_s, x) = 0$$

holds for ***all*** *sensors $i \in \{1, 2, \ldots, N\}$ and* ***all*** *sampled sensor-location-measurement combinations (i, x^i, z^i), then $c_1 = c_2 = \cdots = c_s = 0$.*

Proof. The proof of Theorem V.2 is composed of two steps.

(Step 1). We first prove that the marginalized measurement model $q_k^i(z_k^i|x_k^i) := \sum_\theta p^i(z_k^i|\theta, x_k^i)\prod_{j=1}^N[\pi_{k-1}^j(\theta)]^{P_k[i,j]}$ converges to the true measurement model $p^i(z_k^i|\theta^*, x_k^i)$, w.p.1.

According to Eqn. (5), the estimated probability of the true parameter θ^* by agent i satisfies the following equation,

$$\log \pi_k^i(\theta^*) = \sum_{j=1}^N P_k[i,j]\log \pi_{k-1}^j(\theta^*) + \log\frac{p^i(z_k^i|\theta^*, x_k^i)}{q_k^i(z_k^i|x_k^i)}. \tag{6}$$

Let $\mathcal{F}_k^i = \sigma\left\{(x_t^i, z_t^i), t \leq k\right\}$ be the $\sigma-$algebra [30] generated by the past sampled location-measurement pairs.

Since the collected noisy measurement z_k^i is generated from the true measurement model, i.e., $z_k^i \sim p^i(z_k^i|\theta^*, x_k^i)$, taking expectation with respect to all possible sampled location-measurement pairs on both sides of (6), we have

$$\begin{aligned}\mathbb{E}[\log \pi_k^i(\theta^*)] &= \sum_{j=1}^N P_k[i,j]\mathbb{E}[\log \pi_{k-1}^j(\theta^*)] \\ &\quad + \mathbb{E}\left[\log \frac{p^i(z_k^i|\theta^*, x_k^i)}{q_k^i(z_k^i|x_k^i)}\right] \\ &= \sum_{j=1}^N P_k[i,j]\mathbb{E}[\log \pi_{k-1}^j(\theta^*)] \\ &\quad + \mathbb{E}\left[\mathbb{E}\left[\log \frac{p^i(z_k^i|\theta^*, x_k^i)}{q_k^i(z_k^i|x_k^i)}|x_k^i, \mathcal{F}_{k-1}^{1:N}\right]\right] \\ &= \sum_{j=1}^N P_k[i,j]\mathbb{E}[\log \pi_{k-1}^j(\theta^*)] + \mathbb{E}[d_k^i],\end{aligned}$$

where $d_k^i = D_{\text{KL}}(p^i(z|\theta^*, x_k^i)||q_k^i(z|x_k^i))$ is the *relative entropy* (*Kullback–Leibler divergence*) from $p^i(z|\theta^*, x_k^i)$ to $q_k^i(z|x_k^i)$.

Let $\Pi_k(\theta) = [\mathbb{E}[\log \pi_k^1(\theta)], \mathbb{E}[\log \pi_k^2(\theta)], \dots, \mathbb{E}[\log \pi_k^N(\theta)]]^T$, $Q_k = [\mathbb{E}[d_k^1], \mathbb{E}[d_k^2], \dots, \mathbb{E}[d_k^N]]^T$. Viewing the sensor network as a whole system, then the update rule of the locally estimated posterior distributions can be written in a matrix form:

$$\Pi_k(\theta^*) = P_k\Pi_{k-1}(\theta^*) + Q_k. \tag{7}$$

Expanding (7) iteratively along the time series, we get

$$\begin{aligned}\Pi_k(\theta^*) &= P_k\Pi_{k-1}(\theta^*) + Q_k \\ &= P_k(P_{k-1}\Pi_{k-2}(\theta^*) + Q_{k-1}) + Q_k \\ &= \cdots \\ &= (P_k \cdots P_1)\Pi_0(\theta^*) + \sum_{t=1}^{k-1}[(P_k \cdots P_{t+1})Q_t] + Q_k.\end{aligned} \tag{8}$$

Based on condition (i), according to *Wolfowitz's Lemma* V.1 and Lemma IV.2, there exists a row vector α_t for any $t \in \mathbb{N}$, and $k > t, k \in \mathbb{N}$ such that

$$\lim_{k\to\infty} P_k P_{k-1} \cdots P_{t+2} P_{t+1} = \mathbf{1}\alpha_t.$$

where $\mathbf{1} = [1, \dots, 1]^T$.

Furthermore, since P_t is non-negetive for all $t \in \mathbb{N}$, α_t is also non-negetive for all $t \in \mathbb{N}$. Since $\pi_k^i(\theta^*) \leq 1$, $\log \pi_k^i(\theta^*) \leq 0$ is upper bounded $\forall i \in \{1, 2, \dots, N\}$, which means that each element of $\Pi_k(\theta^*)$ is upper bounded, thus $\Pi_k(\theta^*)$ is also upper bounded. Moreover, since the right-hand side of (8) is non-decreasing due to the fact that the relative entropy d_k^i is non-negative, $\Pi_k(\theta^*)$ is also non-decreasing. Thus, the limit of $\Pi_k(\theta^*)$ exists. Therefore, taking $k \to \infty$ in (8), we obtain

$$\Pi_\infty(\theta^*) = \mathbf{1}\alpha_0\Pi_0(\theta^*) + \sum_{t=1}^\infty(\mathbf{1}\alpha_t Q_t), \tag{9}$$

which indicates that $\Pi_k(\theta^*)$ converges to a vector where each element of the vector is the same.

Then without losing generality, for any certain agent i, its locally estimated probability of the true parameter θ^* satisfies

$$\mathbb{E}[\log \pi_\infty^i(\theta^*)] = \sum_{j=1}^N \alpha_0^j \mathbb{E}[\log \pi_0^j(\theta^*)] + \sum_{t=1}^\infty \sum_{j=1}^N \alpha_t^j \mathbb{E}[d_t^j].$$

As a sequel, we get

$$\begin{aligned}\sum_{t=1}^\infty \sum_{j=1}^N \alpha_t^j \mathbb{E}[d_t^j] &= \mathbb{E}[\log \pi_\infty^i(\theta^*)] - \sum_{j=1}^N \alpha_0^j \mathbb{E}[\log \pi_0^j(\theta^*)] \\ &\leq -\sum_{j=1}^N \alpha_0^j \mathbb{E}[\log \pi_0^j(\theta^*)] < \infty,\end{aligned}$$

where the last inequality holds according to condition (ii) and the fact that α_0 is non-negative.

Therefore, for any $\epsilon > 0$ and $i \in \{1, 2, \dots, N\}$, by *Markov Inequality*, we have

$$\sum_{k=0}^\infty P[d_k^i \geq \epsilon] \leq \frac{1}{\epsilon}\sum_{k=0}^\infty \mathbb{E}[d_k^i] < \infty.$$

We can then apply *Borel-Cantelli Lemma* and show that $P(d_k^i \geq \epsilon, i.o.) = 0$, which further implies $\lim_{k\to\infty} d_k^i = 0, \forall i \in \{1, 2, \dots, N\}$, w.p.1.

Moreover, since $d_k^i \geq 0$, by *Tonelli's Theorem*, we have

$$\mathbb{E}\left[\sum_{k=0}^\infty d_k^i\right] = \sum_{k=0}^\infty \mathbb{E}[d_k^i] < \infty.$$

Then since $\sum_{k=0}^\infty d_k^i$ has bounded expectation, it must be finite w.p.1.

Note that the *total variation distance* between two distributions is related to the *relative entropy* by *Pinsker's Inequality*:

$$||p^i(z|\theta^*, x_k^i) - q_k^i(z|x_k^i)||_{TV} \leq \sqrt{2d_k^i},$$

where

$$||p^i(z|\theta^*, x_k^i) - q_k^i(z|x_k^i)||_{TV} = \sup_z |p^i(z|\theta^*, x_k^i) - q_k^i(z|x_k^i)|.$$

Letting $k \to \infty$, by the convergence of d_k^i, we have

$$\lim_{k\to\infty}\int_z |p^i(z|\theta^*, x_k^i) - q_k^i(z|x_k^i)|dz = 0, w.p.1.$$

According to *Dominated Convergence Theorem*, we further have

$$\int_z \lim_{k\to\infty} |p^i(z|\theta^*, x_k^i) - q_k^i(z|x_k^i)|dz = 0, w.p.1.$$

Moreover, since $|p^i(z|\theta^*, x_k^i) - q_k^i(z|x_k^i)| \geq 0$ and $p^i(z|\theta^*, x_k^i) - q_k^i(z|x_k^i)$ is continuous in z, then for any z,

$$\lim_{k\to\infty} |p^i(z|\theta^*, x_k^i) - q_k^i(z|x_k^i)| = 0, w.p.1,$$

which means

$$\lim_{k\to\infty} q_k^i(z|x_k^i) = p^i(z|\theta^*, x_k^i), w.p.1. \tag{10}$$

(Step 2). We now prove that each agent's locally estimated distribution of the unknown parameter θ, $\pi_k^i(\theta), \forall i \in \{1, 2, \dots, N\}$, all converge to the true distribution $\delta_{\theta^*}(\theta)$, w.p.1.

Let $\hat{\pi}_k^i(\theta) = \prod_{j=1}^N [\pi_{k-1}^j(\theta)]^{P_k[i,j]}$, then we have

$$\begin{aligned}&p^i(z|\theta^*, x_k^i) - q_k^i(z|x_k^i) \\ =&p^i(z|\theta^*, x_k^i) - \sum_\theta p^i(z|\theta, x_k^i)\hat{\pi}_k^i(\theta) \\ =&[1 - \hat{\pi}_k^i(\theta^*)]p^i(z|\theta^*, x_k^i) + \sum_{\theta\neq\theta^*} \hat{\pi}_k^i(\theta)p^i(z|\theta, x_k^i).\end{aligned} \tag{11}$$

Since $\pi_k^i(\theta)$ is bounded for any $i \in \{1, 2, \dots, N\}$, then by *Bolzano–Weierstrass theorem*, there exists a convergent sub-sequence $\{[\pi_k^1(\theta), \pi_k^2(\theta), \dots, \pi_k^N(\theta)]^T, k = t_1, t_2, \cdots\}$, which converges to $[\pi_\infty^1(\theta), \pi_\infty^2(\theta), \dots, \pi_\infty^N(\theta)]^T$.

Moreover, from (9) we know that the expected posterior distributions of each agent all converge to the same limiting distribution. Thus we can take a further sub-sequence $\{\tau_1, \tau_2, \cdots\}$ from $\{t_1, t_2, \cdots\}$ such that the sub-sequence $\{[\pi_k^1(\theta), \pi_k^2(\theta), \ldots, \pi_k^N(\theta)]^T, k = \tau_1, \tau_2, \cdots\}$ converges to $[\pi_\infty(\theta), \pi_\infty(\theta), \ldots, \pi_\infty(\theta)]^T$. As a sequel, the sub-sequence $\{\hat{\pi}_k^i(\theta), k = \tau_1, \tau_2, \cdots\}$ also converges to $\pi_\infty(\theta)$.

Finally, since for any $i \in \{1, 2, \ldots, N\}$, x_k^i is also bounded, we could again take a further sub-sequence $\{m_1, m_2, \cdots\}$ from $\{\tau_1, \tau_2, \cdots\}$ such that $\{x_k^i, k = m_1, m_2, \cdots\}$ converges to x_∞^i.

Now we can take limit over (11) along $\{m_1, m_2, \cdots\}$ and according to the convergence of $q_k^i(z|x_k^i)$, we have

$$[1 - \pi_\infty(\theta^*)]p^i(z|\theta^*, x_\infty) + \sum_{\theta \neq \theta^*} \pi_\infty(\theta) p^i(z|\theta, x_\infty) = 0,$$

for any $i \in \{1, 2, \ldots, N\}$, w.p.1. According to condition (iii), for any convergent sub-sequence, $1 - \pi_\infty(\theta^*) = 0$ and $\pi_\infty(\theta) = 0, \forall \theta \neq \theta^*$, which further implies

$$\pi_\infty(\theta) = \delta_{\theta^*}(\theta), w.p.1.$$

Therefore, for any $i \in \{1, 2, \ldots, N\}$,

$$\lim_{k \to \infty} \pi_k^i(\theta) = \delta_{\theta^*}(\theta), w.p.1. \quad (12)$$

□

Remark 5. When $N = 1$, the cooperative consistency conditions reduce to the consistency conditions for a single agent that the unknown parameter should be identifiable within the sampled dataset, which was stated in our previous work [28], [31]. The cooperative consistency conditions for multi-agent are actually weaker conditions compared to single-agent, since they only require that the unknown parameter be identifiable within the joint sampled dataset for all the agents. For example, there may exist a certain subspace of the parameter space that is not distinguishable for one agent with its locally sampled dataset, and another subspace for another agent. However, by combining the sampled datasets from all the agents, the unknown parameter can become distinguishable within the joint sampled dataset.

The cooperative consistency conditions provide great potential to enhance the network's sensing ability, since different types of sensors can be utilized to collect different types of measurements, which facilitates the richness of the sampled data. Furthermore, since the burden of computation and mobility is distributed to each agent, fewer constraints need to be put on the individual sensor's type and ability.

VI. Simulation

We consider a source of chemical plume, which generates plume particles in the 2D space. The field function is represented by the rate of hits, which is defined as the average number of particles per unit time measured by the sensor at a certain location. The rate of hits for a chemical plume source can be given as:

$$R_\theta(x) = \frac{R_s}{\log \frac{\gamma}{a}} \exp(-\frac{\langle \theta - x, V \rangle}{2D}) K_0(\frac{||\theta - x||_2}{\gamma}), \quad (13)$$

where R_s is the rate at which the plume source releases the plume particles in the environment, $\gamma = \sqrt{D\tau/(1 + \frac{||V||^2 \tau}{4D})}$ is the average distance traveled by a plume particle in its lifetime, a is the size of the sensor detecting plume particles, V is the average wind velocity, D is the diffusivity of the plume particles and K_0 is the Bessel function of zeroth order.

We model the measurement z as a Poisson random variable with $R_\theta(x)\Delta t$ as the rate parameter. The measurement model is then

$$p(z|\theta, x) = \frac{\exp(-R_\theta(x)\Delta t)(R_\theta(x)\Delta t)^z}{z!}.$$

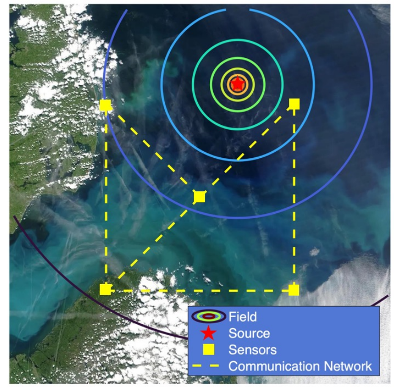

Fig. 1: Graphical representation of the simulation scenario. A multi-sensor network that consists of 5 sensors with periodically connected communication is deployed to estimate the source location in a stochastic source field.

The performance of the proposed algorithm is assessed by numerical simulations. We perform the simulation for a model of chemical plume where detectable particles are emitted at rate $R = 10$, have a lifetime $\tau = 2500$, propagate with diffusivity $D = 10$, and $V = [0, 0]$ in the absence of wind. We assume the sensor size $a = 1$, and the sensor takes $\Delta t = 5s$ to take a measurement of the signal. We consider a multi-sensor network that consists of 5 sensors and the dynamic communication network is periodically connected, as illustrated in Fig. 1. The time-varying communication topology has a period of 3: $t_{3k+1} : \{1 \leftrightarrow 2, 1 \leftrightarrow 4, 3 \leftrightarrow 5\}, t_{3k+2} : \{1 \leftrightarrow 2, 3 \leftrightarrow 5, 4 \leftrightarrow 5\}, t_{3k+3} : \{1 \leftrightarrow 2, 1 \leftrightarrow 3, 2 \leftrightarrow 3, 4 \leftrightarrow 5\}, k \in \mathbb{N}$. The adjacent matrix $A_t[i, j] = 1$ if i and j are connected at time t. The discrete step size of the Perron matrix is chosen as $\epsilon = 0.5$.

We compare the performance of our proposed distributed Bayesian data fusion algorithm with other data fusion methods in both distributed and centralized conditions. For the centralized condition, we compare it with the centralized Bayesian data fusion algorithm, as well as a consensus-based non-Bayesian data fusion algorithm that simply merges all local estimates together with the aim of achieving consensus. For the distributed condition, we compare it with another classic distributed data fusion approach, i.e., the naive averaging approach, which combines all neighboring estimations by calculating their algorithmic average. As demonstrated in Fig. 2, with periodically connected communication, the proposed algorithm converges at a slower rate than the centralized Bayesian algorithm due to the incomplete communication of the sensor network and the conservative WEP fusion rule. The convergence rate of the distributed algorithm may depend on the communication topology, as well as the choice of the adjacent matrix and the discrete step size ϵ. However, the estimation of the distributed algorithm is smoother and more robust than the centralized Bayesian algorithm, since the propagation of the information through the conservative fusion rule naturally serves as a function of filtering and smooths the noises of individual measurements. Furthermore, the proposed method significantly outperforms in terms of convergence rate than other non-Bayesian data fusion approaches in both distributed and centralized conditions. It is also worth noting that the distributed naive averaging approach demonstrates evident instability during the process. This may be due to its incapacity to consistently deal with the common information, which further justified the advantages of the proposed approach.

To validate the cooperative consistency conditions, we also perform the simulation for the same sensor network but without communication. Since each sensor is static, the consistency conditions for every single agent cannot be satisfied. Therefore, the unknown parameter is not identifiable for every single agent, which is demonstrated in Fig. 2 that the TV distance of each local estimate cannot converge without communication. However, when there exists periodically connected communication, the weaker cooperative consistency conditions can be satisfied despite the fact that the consistency conditions for each individual agent are not satisfied. The demand for sufficient data is distributed to multiple agents and the unknown parameter is now distinguishable in the joint sample space.

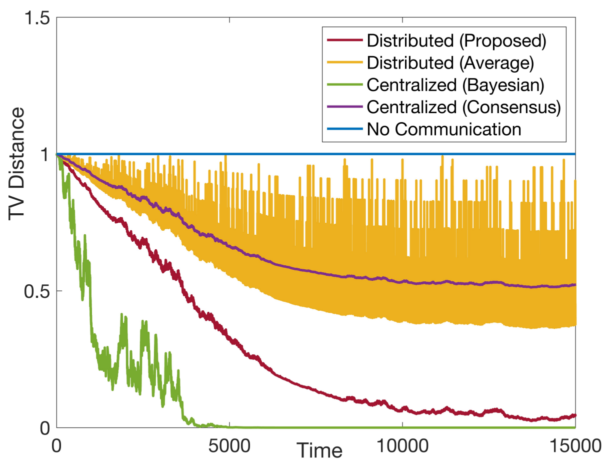

Fig. 2: Comparison of Total Variation (TV) distance between the estimated posterior distribution and the true distribution of the source location for both distributed and centralized data fusion algorithms.

VII. Conclusion and Future Work

In this paper, we presented a uniformly consistent distributed Bayesian data fusion algorithm and rigorously derived the cooperative consistency conditions that guarantee uniform consistency. The performance and convergence properties of the proposed algorithm are validated through simulations. Future works will explore different strategies for determining the trust matrix of the communication network and analyze their convergence rates.

References

[1] Mark E. Campbell and Nisar R. Ahmed. Distributed Data Fusion: Neighbors, Rumors, and the Art of Collective Knowledge. *IEEE Control Systems Magazine*, 36(4):83–109, 8 2016.

[2] Geoffrey A. Hollinger, Srinivas Yerramalli, Sanjiv Singh, Urbashi Mitra, and Gaurav S. Sukhatme. Distributed Data Fusion for Multirobot Search. *IEEE Transactions on Robotics*, 31(1):55–66, 2 2015.

[3] Asif Khan, Bernhard Rinner, and Andrea Cavallaro. Cooperative robots to observe moving targets: Review. *IEEE Transactions on Cybernetics*, 48(1):187–198, 1 2018.

[4] Jun Chen and Philip Dames. Distributed Multi-Target Tracking for Heterogeneous Mobile Sensing Networks with Limited Field of Views. In *2021 IEEE International Conference on Robotics and Automation (ICRA)*, pages 9058–9064. IEEE, 5 2021.

[5] Dongdong Yu, Yuanqing Xia, Di-Hua Zhai, and Yufeng Zhan. On distributed fusion estimation with stochastic scheduling over sensor networks. *Automatica*, 142:110406, 8 2022.

[6] Federico Castanedo. A Review of Data Fusion Techniques. *The Scientific World Journal*, 2013:1–19, 2013.

[7] Hugh Durrant-Whyte and Thomas C. Henderson. Multisensor Data Fusion. In *Springer Handbook of Robotics*, volume 35, pages 867–896. 2016.

[8] Rajnish Kumar, Matthew Wolenetz, Bikash Agarwalla, JunSuk Shin, Phillip Hutto, Arnab Paul, and Umakishore Ramachandran. DFuse: A Framework for Distributed Data Fusion. In *Proceedings of the 1st international conference on Embedded networked sensor systems*, pages 114–125, New York, NY, USA, 11 2003. ACM.

[9] Qinyuan Liu, Zidong Wang, Xiao He, and Donghua Zhou. Event-Based Distributed Filtering Over Markovian Switching Topologies. *IEEE Transactions on Automatic Control*, 64(4), 2019.

[10] Xingkang He, Wenchao Xue, Xiaocheng Zhang, and Haitao Fang. Distributed filtering for uncertain systems under switching sensor networks and quantized communications. *Automatica*, 114:108842, 4 2020.

[11] Jiahao Zhang, Shesheng Gao, Guo Li, Juan Xia, Xiaomin Qi, and Bingbing Gao. Distributed recursive filtering for multi-sensor networked systems with multi-step sensor delays, missing measurements and correlated noise. *Signal Processing*, 181:107868, 4 2021.

[12] Ronald P. S. Mahler. Optimal/robust distributed data fusion: a unified approach. In *Signal Processing, Sensor Fusion, and Target Recognition IX*, volume 4052, pages 128–138. SPIE, 8 2000.

[13] Stefano Battilotti and Filippo Cacace. A stability with optimality analysis of consensus-based distributed filters for discrete-time linear systems. *Automatica*, 129:109589, 2021.

[14] Jose Cadena, Priyadip Ray, Hao Chen, Braden Soper, Deepak Rajan, Anton Yen, and Ryan Goldhahn. Stochastic gradient-based distributed bayesian estimation in cooperative sensor networks. *IEEE Transactions on Signal Processing*, 69:1713–1724, 2021.

[15] All Jadbabaie, Jie Lin, and A. Stephen Morse. Coordination of groups of mobile autonomous agents using nearest neighbor rules. *IEEE Transactions on Automatic Control*, 48(6):988–1001, 6 2003.

[16] Wei Ren and Randal W. Beard. Consensus seeking in multiagent systems under dynamically changing interaction topologies. *IEEE Transactions on Automatic Control*, 50(5):655–661, 5 2005.

[17] Luc Moreau. Stability of multiagent systems with time-dependent communication links. *IEEE Transactions on Automatic Control*, 50(2):169–182, 2 2005.

[18] Nisar Ahmed, Jonathan Schoenberg, and Mark Campbell. Fast Weighted Exponential Product Rules for Robust General Multi-Robot Data Fusion. In *Robotics: Science and Systems*, pages 9–16, 2013.

[19] Giorgio Battistelli and Luigi Chisci. Kullback–Leibler average, consensus on probability densities, and distributed state estimation with guaranteed stability. *Automatica*, 50(3):707–718, 3 2014.

[20] Saptarshi Bandyopadhyay and Soon Jo Chung. Distributed Bayesian filtering using logarithmic opinion pool for dynamic sensor networks. *Automatica*, 97:7–17, 11 2018.

[21] Günther Koliander, Yousef El-Laham, Petar M Djuri´c, and Franz Hlawatsch. Fusion of Probability Density Functions. *Proceedings of the IEEE*, 110, 2022.

[22] Simon J. Julier, Tim Bailey, and Jeffrey K. Uhlmann. Using Exponential Mixture Models for Suboptimal Distributed Data Fusion. In *2006 IEEE Nonlinear Statistical Signal Processing Workshop*, pages 160–163. IEEE, 9 2006.

[23] Simon J Julier. Estimating and exploiting the degree of independent information in distributed data fusion. In *2009 12th International Conference on Information Fusion*, pages 772–779, 2009.

[24] D. Franken and A. Hupper. Improved fast covariance intersection for distributed data fusion. In *2005 7th International Conference on Information Fusion*, pages 154–160, 2005.

[25] Jeffrey K. Uhlmann. Covariance consistency methods for fault-tolerant distributed data fusion. *Information Fusion*, 4(3):201–215, 9 2003.

[26] K. C. Chang, Chee Yee Chong, and Shozo Mori. Analytical and computational evaluation of scalable distributed fusion algorithms. *IEEE Transactions on Aerospace and Electronic Systems*, 46(4):2022–2034, 10 2010.

[27] Reza Olfati-Saber, J. Alex Fax, and Richard M. Murray. Consensus and Cooperation in Networked Multi-Agent Systems. *Proceedings of the IEEE*, 95(1):215–233, 1 2007.

[28] Yingke Li, Tianyi Liu, Enlu Zhou, and Fumin Zhang. Bayesian Learning Model Predictive Control for Process-Aware Source Seeking. *IEEE Control Systems Letters*, 6:692–697, 2022.

[29] J. Wolfowitz. Products of Indecomposable, Aperiodic, Stochastic Matrices. *Proceedings of the American Mathematical Society*, 14(5):733, 10 1963.

[30] Achim Klenke. *Probability Theory: A Comprehensive Course*. 2014.

[31] Tianyi Liu, Yifan Lin, and Enlu Zhou. Bayesian Stochastic Gradient Descent for Stochastic Optimization with Streaming Input Data. *SIAM Journal on Optimization*, 34(1):389–418, 2024.