2019 IEEE 58th Conference on Decision and Control (CDC)
Palais des Congrès et des Expositions Nice Acropolis
Nice, France, December 11-13, 2019

Hypothesis assignment and partial likelihood averaging for cooperative estimation

Parth Paritosh, Nikolay Atanasov and Sonia Martinez

Abstract— **We propose a cooperative, decentralized inference algorithm allowing sensor networks to learn a joint parameter best explaining their combined observations. This joint parameter is represented via a probability density over a discrete set of hypotheses. We aim to answer two questions: (i) an agent-hypothesis assignment problem, balancing estimation quality, storage and communication constraints in the networks, and (ii) the design of a provably-correct distributed estimation algorithm on limited hypothesis sets for agents. We make the following contributions to the state of the art. First, our proposed algorithm allows each agent to perform updates on partial likelihoods and exchange local information on a limited hypothesis set, as opposed to the entire hypothesis space. For some of the agents, the limited hypothesis domains may even exclude the true hypothesis. Second, the presented algorithm is the first to not require step-wise renormalization at all hypotheses, while still guaranteeing consensus and convergence of sensor estimates. Third, we also address agent-hypothesis assignment by formulating it as a mixed integer programming problem, that matches agent sub-networks to hypotheses based on a diversity criterion for estimation quality. We provide numerical examples demonstrating the benefits of these algorithms.**

I. Introduction

Sensor networks have been widely deployed for in situ data gathering and environment monitoring, enabling the estimation of relevant parametric models from data. In these settings, a distributed estimation process overcomes physical limitations in the communication between interconnected systems [1], [2], any single-point failures, and privacy concerns on data sharing [3]. A main tool to solve cooperative estimation problems online is based on (consensus) non-Bayesian learning algorithms, which are protocols governed by the recursive interactions of single-hop neighbors. Typically, these algorithms require agents to exchange information on large hypothesis sets, which vastly increases the communication and storage cost for large sensor networks and environments. Motivated by the cost issue, we look at the design of more scalable algorithms that rely on partial likelihood updates. We have assumed that sensors receive source measurements infinitely often to infer source state on a finite discrete space.

Literature review: Distributed consensus algorithms have been designed continually since the early 70's. The initial studies were aimed at developing a Bayesian framework for agreement between two individual sets of information [4]. In the following decade, there were results published on the effect of network topology on averaging opinions among multiple agents [5], [6]. Other aspects such as network error and bit rate constraints have also been considered in the past decade [7]. A major improvement appeared in the form of non-Bayesian updates [8], performed by updating the hyper-parameters of a probability density function (pdf) instead of the updating the function itself. The stationary distribution and convergence rates of this approach have recently been studied in [9], [10]. Even though recent approaches can deal with network-level updates, they require maintaining and communicating each agent's pdf over the entire set of hypotheses. Drawing inspiration from the idea of distributed computation, it is practically useful to consider a distributed storage scheme, in which agents only maintain and exchange a partial pdf over the parameter domain.

Partial likelihood updates are predicated on assigning sets of hypotheses to individual agents. The agent-hypothesis assignment problem has its roots in classical matching problems [11]. More recently it has been employed for sensor network assignment [12]. The assignment problems have usually been tackled via integer-programming or their convex relaxations [13]. Since we deal with optimal sub-networks, the relevant recent works include maximum-weight connected subgraph problems [14], [15] for connected sensor sub-network selection. While the methodology suffices for learning a single subgraph, it is insufficient for finding multiple subgraphs coupled with cardinality constraints on selected agents as needed for the assignment problem presented here.

Statement of Contributions: In this manuscript we address two complementary problems relevant to the distributed inference of a joint parameter by a sensor network. First, we formulate a novel hypothesis assignment problem and propose mixed integer programming solution, for matching subsets of connected agents to different hypothesis sets with the goal of providing good complementary observations while respecting storage and communication constraints of agents. Second, we propose and analyze convergence of a cooperative estimation algorithm which, not only is distributed across one-hop neighborhoods, but also allows agents to maintain likelihoods over a subset of the hypothesis space. Thus, it significantly reduces storage and communication costs in comparison to existing approaches. This algorithm is free of any network-wide normalization updates present in existing consensus algorithms based on geometric averaging.

II. Problem Formulation

We consider a set of sensors $\mathcal{N} = \{1, \dots, n\}$ whose communications are modeled via an undirected graph $\mathcal{G} =$

We gratefully acknowledge support from NSF NRI CNS-1830399, ONR N00014-19-1-2471, DARPA Lagrange N660011824027, AFOSR FA9550-18-1-0158 and ARL DCIST CRA W911NF-17-2-0181.

The authors are with Contextual Robotics Institute, University of California San Diego, 9500 Gilman Dr, La Jolla, CA 92093 {pparitos,natanasov,soniamd}@ucsd.edu

$(\mathcal{N}, \mathcal{E})$, with edges $\mathcal{E}$ representing the communicating pairs. Each agent $i \in \mathcal{N}$ has a corresponding state variable $\boldsymbol{x}_i \in \mathbb{R}^{d_x}$ and receives data $\boldsymbol{z}_{i,t} \in \mathbb{R}^{d_z}$ at time t. Based on sensor i's characteristics and the state variables $\boldsymbol{x}_i$, its observation model is specified by a pdf $\mathrm{pz}_i(\boldsymbol{z}|\boldsymbol{\theta})$ defined on discrete domain Θ_i with cardinality m_i. The network aims to find the true value of a joint parameter $\boldsymbol{\theta} \in \mathbb{R}^{d_\theta}$ in the finite discrete set $\Theta \equiv \cup_{i\in\mathcal{N}}\Theta_i$, which may represent the possible locations of a data-generating source. This set Θ is discrete and finite with cardinality $|\Theta| = m$. The true set of parameter values generating the agent observations is $\Theta^* \subset \Theta$. The set of neighbors for each agent i in the communication graph $\mathcal{G}$ is defined as $\mathcal{N}_i$. Each agent i maintains the probability $p_{i,t}(\boldsymbol{\theta}), \boldsymbol{\theta} \in \Theta_i$, denoting agent's confidence on the correctness of hypothesis $\boldsymbol{\theta}$ at time step t. The agents use their observations $\boldsymbol{z}_{i,t}$ and neighbor estimates $p_{j,t}(\boldsymbol{\theta})$, $(i,j) \in \mathcal{E}$ at hypothesis $\boldsymbol{\theta}$ to update their own probability density $p_{i,t+1}(\boldsymbol{\theta})$ over the values of $\boldsymbol{\theta} \in \Theta_i$.

Since agent i tracks and shares inference only over the subset Θ_i, the agent's computational and communications load is diminished. The coverage of entire hypothesis space is ensured with the condition $\cup_{i\in\mathcal{N}}\Theta_i = \Theta$, in effect assigning at least one agent for tracking every hypothesis. We denote the set of agents observing a specific hypothesis v, $\boldsymbol{\theta}_v \in \Theta$, as $\mathcal{V}(\boldsymbol{\theta}_v) \subseteq \mathcal{N}$. If the agents in $\mathcal{V}(\boldsymbol{\theta}_v)$ form a connected subgraph $\mathcal{G}_{\theta_v} \equiv \mathcal{G}_v$ of $\mathcal{G}$, we assign a corresponding doubly stochastic communication matrix $\boldsymbol{A}(\boldsymbol{\theta}_v) \equiv \boldsymbol{A}_v$ to the hypothesis $\boldsymbol{\theta}_v$. The matrix $\boldsymbol{A}_v$ is induced from $\mathcal{G}$, following the same sign pattern as the adjacency matrix of the original graph. The doubly stochastic matrix is selected for its averaging properties [16] that we use to prove asymptotic convergence in Section VI. We aim to address the following questions dealing with the computation and communication complexity for distributed estimation.

Hypothesis assignment. How should the parameter hypotheses be distributed among the agents to balance storage and communication requirements with estimation quality? We provide details on the following four modeling criteria to achieve these requirements,

a) Agent diversity: The diversity requirement assigns a set of agents $\mathcal{V}(\boldsymbol{\theta})$ with diverse observation models at each hypothesis $\boldsymbol{\theta}$ with the goal of augmenting quality of inferences from observations. This can be realized by ensuring that the assigned set of agents $\mathcal{V}(\boldsymbol{\theta})$ maximize the symmetric KL divergence metric $\mathrm{D_{KL}}$ measuring distance between among pairs of hypothesis pdf $\mathrm{pz}_i(\cdot|\boldsymbol{\theta}), i \in \mathcal{V}(\boldsymbol{\theta})$.

$$F(\mathcal{G}_{\boldsymbol{\theta}}) = \sum_{\substack{i,j\in\mathcal{V}(\boldsymbol{\theta})\\(i,j)\in\mathcal{E}}} \mathrm{D_{KL}}(\mathrm{pz}_i(\cdot|\boldsymbol{\theta}), \mathrm{pz}_j(\cdot|\boldsymbol{\theta})) + \sum_{i\in\mathcal{V}(\boldsymbol{\theta})} \mathrm{H}(\mathrm{pz}_i(\cdot|\boldsymbol{\theta})). \tag{1}$$

We induce the diversity criterion with the KL divergence metric $\mathrm{D_{KL}}$ while the entropy H term favors agents with flatter observation densities pz_i. It is worth noting that the divergence factor depends on the edges $(i,j) \in \mathcal{E}$ capturing the synergies of one-hop neighbors.

b) Connectivity: This requires that the subset of agents $\mathcal{V}(\boldsymbol{\theta}) \in \mathcal{N}$ assigned to each hypothesis $\boldsymbol{\theta}$ is connected. The connected sets of agents learn from their neighbors, thus leading to consistent estimates.

c) Computational load: This is implemented by limiting the cardinality on the hypotheses observed by individual agents, $0 < |\Theta_i| \leq m_i$. This constraint caps the number of hypotheses tracked by agents at every time step, making the algorithm scalable in storage and communication.

d) Coverage: This translates into the requirement $\cup_i\Theta_i = \Theta$, ensuring that every hypothesis is being tracked by at least one agent.

Thus, our first research question is addressed by means of an optimization problem over sets of subgraphs $\mathcal{G}_{\boldsymbol{\theta}}, \forall\boldsymbol{\theta} \in \Theta$,

$$\max_{\{\mathcal{G}_{\boldsymbol{\theta}}\}_{\boldsymbol{\theta}\in\Theta}} \quad \sum_{\boldsymbol{\theta}\in\Theta} F(\mathcal{G}_{\boldsymbol{\theta}}), \tag{2}$$

$$\mathcal{G}_{\boldsymbol{\theta}} \text{ is a connected induced subgraph of } \mathcal{N},\ \forall\boldsymbol{\theta} \in \Theta, \tag{3}$$

$$0 < |\Theta_i| \leq m_i\ \forall i \in \mathcal{N}, \tag{4}$$

$$\cup_i\Theta_i = \Theta. \tag{5}$$

In Section III, we develop an approach to solve this problem of optimal hypothesis assignment offline.

Distributed inference on limited hypothesis sets. How can communicating agents merge their observations on a limited set of hypotheses to achieve a consistent inference on the probability density of true source parameter? To answer this question, we aim to develop a distributed learning algorithm by which agents merge their partial likelihoods with neighboring agents' beliefs to find the true hypothesis. That is, through such algorithm agents will merge their estimates only on their assigned subset $\Theta_i \subset \Theta$, and not over the complete Θ. To arrive at a network-wide consensus on true hypothesis at time T, each agent i will make use of data $\boldsymbol{z}_{i,1:T}$ relative to Θ_i. Thus, each agent is tasked with arriving at the values of a probability mass function $p_{i,T}(\boldsymbol{\theta})$ over only the hypotheses $\boldsymbol{\theta} \in \Theta_i$ at time T using neighbor estimates and data $\boldsymbol{z}_{i,1:T}$ (for the values $\boldsymbol{\theta} \in \Theta \setminus \Theta_i$ at time T the agent will collectively assign a complementary mass value, $\sum_{\boldsymbol{\theta}\in\Theta\setminus\Theta_i} p_{i,T}(\boldsymbol{\theta}) = 1 - \sum_{\boldsymbol{\theta}\in\Theta_i} p_{i,T}(\boldsymbol{\theta})$). As $T \to \infty$, the algorithm should converge to a common distribution $p_\infty(\boldsymbol{\theta})$ that assigns mass to only the true hypothesis set $\Theta^* = \arg\max_{\boldsymbol{\theta}} p(\boldsymbol{\theta}|\boldsymbol{z}_{1:n,1:T})$ and which best explains the observations of all agents. To achieve consistent estimates, the probabilities estimated by agents at any shared hypothesis should converge to the same value i.e. $\lim_{t\to\infty} p_{i,t}(\boldsymbol{\theta}) = p_\infty(\boldsymbol{\theta})$, $\forall i \in \{1,\ldots,n\}$. Further, $p_\infty(\boldsymbol{\theta}) = 0$, $\forall\boldsymbol{\theta} \notin \Theta^*$.

In Section IV, we present a distributed consensus algorithm for agents to merge inferences using partial space likelihood models.

III. Defining the agent network

This section addresses the Hypothesis Assignment question of Section II. We start by analyzing the properties of the objective function encoding diversity as defined in Eqn. (2). We have mentioned existing solutions to relaxed versions of the assignment problem, followed with a novel mixed integer programming formulation to assign agent subgraphs to each hypothesis and simulations illustrating its performance on small and large scale networks.

A. Properties of the optimization problem

We first show that the diversity function satisfies the property of increasing marginal returns, or *supermodularity*. Supermodularity on discrete spaces is analogous to concavity in continuous spaces.

Definition 1 (Set function characteristics)**.** Consider a non-negative function $F : 2^X \to \mathbb{R}_{\geq 0}$ defined over a discrete set X. Let $X_1, X_2 \subseteq X$. The function F is supermodular if $F(X_1 \cap X_2) + F(X_1 \cup X_2) \geq F(X_1) + F(X_2)$.

Proposition 1. *The diversity function F in Eqn. (1) is supermodular over the input subgraph set.*

Proof. With node sets $X_1 \subseteq X_2 \subseteq \mathcal{N}$ with element $j \in \mathcal{N} \setminus X_2$, we can state an equivalent definition of supermodularity,

$$F(X_2 \cup \{j\}) - F(X_2) \geq F(X_1 \cup \{j\}) - F(X_1). \quad (6)$$

The supermodularity of objective function defined in Eqn. (1) follows trivially from this definition. □

We provide examples of the connectivity and cardinality requirements on assignments for hypotheses and agents, respectively. Consider a set of agents $\{a, b, c, d\}$ and hypotheses $\{1, \ldots, 9\}$, illustrated in Fig. (1). An example of the cardinality constraint in Eqn. (4) is that agent a does not track more than 4 hypotheses, e.g. $\{1, 3, 5, 9\}$. An example of the connectivity requirement is hypothesis 2 being tracked by a connected subgraph consisting of $\{b, c, d\}$.

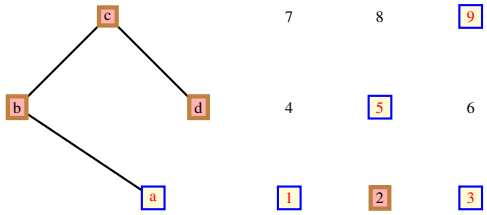

Fig. 1. Group of agents (left) performing inference over a set of hypotheses (right). Cardinality constraint on a as $|\Theta_a| \leq 4$. Connectivity constraints on hypothesis 2 requires corresponding observing agents to be connected $\{b, c, d\}$ in communication graph.

Subgraph selection for hypotheses: Next, we study how to select connected subgraphs from the original network for maximizing diversity function under cardinality constraints imposed on hypothesis sets at each node. First, it is clear that a feasible solution exists if $\sum_{i \in \mathcal{N}} m_i \geq |\Theta|$. Even though the optimization problem in Eqn. (2-5) has a supermodular objective function, the agent subset connectivity constraints in Eqn. (3) render the optimization NP-hard.

There are existing solutions in the literature for two relaxed versions of the optimization problem in Eqn. (2). Relaxing the connectivity constraints, the problem simplifies to the maximization of a supermodular objective under cardinality constraints. It can be solved either as an integer linear program or via Lovász's extension [17] of supermodular objective function in continuous space via polynomial time algorithms [18]. Whereas if we relax the cardinality and coverage constraints while keeping the connected subgraph constraints, the network level optimization problem becomes a hypothesis separable problem. It is same as the generalized maximum-weight connected subgraph problem [15]. A simpler version of this problem was first mentioned in the list of NP-complete problems [19].

B. Proposed optimization strategy

To solve the original problem in Eqn. (2-5), we first formulate the problem by relaxing the connectivity constraint via binary variables representing inclusion of nodes and edges in the optimal subgraph. The mixed integer programming approach presented here is a novel formulation for finding optimal connected subgraphs with cardinality constraints limiting the count on selection of nodes across optimal subgraphs. Consider $\mathbf{y}_v = [y_{1v}, \ldots, y_{nv}]^\top \equiv [y_{iv}]_{i=1}^n \in \{0,1\}^m$, where $y_{iv} = 1$ implies that agent i tracks probabilities for hypothesis v. Consider another variable $\mathbf{b}_v = [b_{ij_1,v}, \ldots b_{ij_\ell,v}] \in \{0,1\}^\ell$, where with a slight abuse of notation we denote one of the $\ell = |\mathcal{E}|$ edges in the agent network as $ij_l \equiv (i,j)_l \in \mathcal{E}$, $l \in \{1, \ldots, \ell\}$, and we use the shorthand $b_{ij,v}$ to refer to a generic entry of $\mathbf{b}_v$. Now, exploiting the separability of the problem, we focus on the reformulation of the assignment of agents to each single hypothesis $\boldsymbol{\theta}_v$, as explained next.

The objective function can be written in terms of the KL-divergence between communicating agents. This function is linear in terms of the new node and edge binary variables. The cardinality constraint for agent i is expressed in terms of the nodal variables y_{iv}. The coverage constraint is satisfied if each hypothesis $\boldsymbol{\theta}_v$ is observed by at least one agent. This results into the following linear program,

$$\sum_{v=1}^{m} \Big[\max_{\mathbf{y}_v, \mathbf{b}_v} \sum_{(i,j) \in \mathcal{E}} b_{ij,v} \, \mathrm{D_{KL}}(\mathrm{pz}_i(z|\boldsymbol{\theta}_v), \mathrm{pz}_j(z|\boldsymbol{\theta}_v)) + \sum_{i=1}^{n} y_{iv} \, \mathrm{H}(\mathrm{pz}_i(z|\boldsymbol{\theta}_v)) \Big] \quad (7)$$

$$\sum_{v=1}^{m} y_{iv} \leq m_i, \quad \forall i \in \{1, \ldots, n\}, \qquad \text{(Cardinality)}$$

$$\sum_{i=1}^{n} y_{iv} \geq 1, \quad \forall v \in \{1, \ldots, m\}. \qquad \text{(Coverage)}$$

Now, we can enforce graph connectivity on trees by adding a set of constraints on the edges as in Eqn. (8-9). From Eqn. (9), we can note that nodes i, j defining a selected edge (i,j) with $b_{ij,v} = 1$ are automatically selected, i.e. $y_{iv}, y_{jv} = 1$. If $\sum_{i=1}^{n} y_{iv} = 2$, then $\sum_{ij \in \mathcal{E}} b_{ij,v} = 1$ implying selection of exactly one edge and the two nodes defining it.

$$\sum_{ij \in \mathcal{E}} b_{ij,v} = \sum_{i=1}^{n} y_{iv} - 1, \; \forall v \in \{1, \ldots, m\}, \quad (8)$$

$$b_{ij,v} \leq y_{iv}, y_{jv}, \; \forall (i,j) \in \mathcal{E}, \; \forall v \in \{1, \ldots, m\}. \quad (9)$$

Proposition 2. *In an acyclic connected graph, the constraints connecting node and edge variables in Eqn.* (8) *and* (9) *lead to the selection of a connected subgraph.*

Proof. The difference between number of selected nodes and edges can not be larger than one due to Eqn. (8). And since the constraints will yield some collection of connected trees of $\mathcal{G}$ each with difference between edges and nodes being one, it leads to selection of one connected acyclic graph at each hypothesis. □

Based on Proposition (2), a possible algorithm to assign agents to hypotheses consists of first obtaining a spanning tree out of the connected graph and then solving the assignment linear programming problem with constraints given in Eqn. (7-9). Other than this, for graphs containing cycles, there are two alternative ways to enforce connectivity, namely average node degree and a,b-separation [20].

Average node degree: The average node degree of a network is defined as the average over all node degrees in the network. If each edge assigns a value of 1 to its corresponding nodes, then as per [21], the maximum average degree of a tree of k vertices is $2 - 2/k$. If there is a cycle in the graph, the maximum average degree is ≥ 2. Therefore, we can introduce flow variables $f^i_{ij}, f^j_{ij} \in \mathbb{R}_{\geq 0}$ on each edge $(i,j) \in \mathcal{E}$ with $f^i_{ij} + f^j_{ij} = 2$, and add a constraint on the average degree of the optimal subgraph to guarantee connectivity. However, the average node degree method introduces quadratic constraints. The flow variables are expressed for each hypothesis $\boldsymbol{\theta}_v$ in a similar fashion to $\boldsymbol{b}_v$ as $\mathbf{f}_v = [f^i_{ij_1,v}, f^j_{ij_1,v}, \dots f^i_{ij_\ell,v}, f^j_{ij_\ell,v}] \in \mathbb{R}^{2\ell}_{\geq 0}$ with $\ell = |\mathcal{E}|$. Therefore, these constraints are expressed as

$$f^i_{ij,v} + f^j_{ij,v} = 2, \forall (i,j) \in \mathcal{E}, \boldsymbol{\theta}_v \in \Theta \tag{10}$$
$$\sum_{j \in \mathcal{N}_i} f^j_{ij,v} \leq 2 - \frac{2}{\sum_{i=1}^n y_{iv}} \forall i \in \mathcal{N}, \boldsymbol{\theta}_v \in \Theta$$

The constraints in Eqn. (10) can be expressed as a linear and quadratic constraint. In our **solution**, these are used in conjunction with Eqn. (7- 9) to find the optimal graph structures. One can employ *a,b-separation* presented in [20] for finding optimal connected graphs instead of trees. This formulation introduces n^2ml variables in comparison to $2ml$ variables in the formulation based on average node degree in the network.

C. Defining hypothesis-specific communication matrices

The hypothesis assignment optimization matches each hypothesis $\boldsymbol{\theta}_v$ to a set of connected agents $\mathcal{V}(\boldsymbol{\theta}_v)$. A doubly stochastic communication matrix $\boldsymbol{A}(\boldsymbol{\theta}_v) \equiv \boldsymbol{A}_v$ is assigned to the induced subgraph $\mathcal{G}_{\boldsymbol{\theta}_v}$. One popular method to generate doubly stochastic matrices is via iterative normalization along rows and columns of a matrix respecting the underlying connectivity [22]. Let the number of agents tracking probability at hypothesis $\boldsymbol{\theta}_v$ be $n_v = |\mathcal{V}(\boldsymbol{\theta}_v)|$. Also, define a vector of ones as $\mathbf{1}_{n_v} \in \mathbb{R}^{n_v}$. As per [16], the ergodicity of the assigned doubly stochastic matrix $A(\boldsymbol{\theta}_v)$ with positive diagonal elements ensures that $\lim_{t\to\infty} \boldsymbol{A}(\boldsymbol{\theta}_v)^t = \frac{1}{n_v}\mathbf{1}_{n_v}\mathbf{1}^\top_{n_v}$. This introduces the averaging properties needed for designing the likelihood averaging algorithm in the next section.

IV. Distributed consensus on partial hypotheses

In this section, we propose and analyze a network-wide inference algorithm that can be performed with partial observation likelihoods for each agent. The main property of this algorithm is that it does not require group-wide renormalization, yet it has performance guarantees. This lack of renormalization allows for a distributed and more efficient implementation of the algorithm.

For this section, we assume that sets of hypotheses Θ_i tracked by agent i are computed offline. The sets can be computed with the approach presented in Section III. Each agent i thus knows the weights $\{\boldsymbol{A}(\boldsymbol{\theta}_v)_{ij} | \forall j \in \mathcal{V}(\boldsymbol{\theta}_v) \cup \{i\}\}$ placed on beliefs at each $\boldsymbol{\theta}_v \in \Theta_i$. As stated in Section II, each agent aims to reach a consensus on the probability distribution on every hypothesis $\lim_{t\to\infty} p_{i,t}(\boldsymbol{\theta}) = p_\infty(\boldsymbol{\theta}), \boldsymbol{\theta} \in \Theta_i, \forall i \in \{1, \dots, n\}$.

In the state of the art, distributed estimation algorithms are an analogue to Bayesian updates. This type of learning rule can be seen in [10]. The rule can be decomposed in three steps. At each hypothesis, the agent first performs opinion pooling via geometric averaging of its neighbor probabilities as a prior followed by a product with the agent's observation likelihood. The third step is normalization w.r.t. the sum of estimate across all hypotheses Θ as observed by the network. Selecting the weights at the opinion pooling step to form a row or doubly stochastic matrix $\boldsymbol{A}$ ensures convergence to the expectation of network wide weighted average and exact average, respectively [16]. That is in [10],

$$p_{i,t+1}(\boldsymbol{\theta}) = \frac{1}{Z_{i,t+1}} p_i(\boldsymbol{z}_{i,t+1}|\boldsymbol{\theta}) \prod_{j=1}^n p_{j,t}(\boldsymbol{\theta})^{[A]_{ij}}, \ \forall \boldsymbol{\theta} \in \Theta,$$
$$Z_{i,t+1} = \sum_{\boldsymbol{\theta} \in \Theta} \left(p_i(\boldsymbol{z}_{i,t+1}|\boldsymbol{\theta}) \prod_{j=1}^n p_{j,t}(\boldsymbol{\theta})^{[A]_{ij}} \right). \tag{11}$$

We note that, in this algorithm, the updates are defined on the entire hypotheses space Θ and that each agent computes its normalization factor at every time step. Directly using the algorithm in [10] over partial space Θ_i would require extensive bookkeeping, making the step extremely costly to perform. Therefore, we develop a methodology to perform this computation without computing the normalization factor.

A. Partial likelihood estimation algorithm

We propose Algorithm 1 for probability updates with partial likelihoods. The updates for an agent i require the data received at each time, neighbor estimates on assigned hypotheses and stochastic weights specifying their interaction with neighbors at each hypothesis. The algorithm performs the first two steps of opinion pooling and likelihood product, but delays normalization until the very final time step T. In this sense, we can characterize the proposed algorithm as 'Update then Normalize' type in contrast to existing 'Update and Normalize' updates. The lack of normalization step in intermediate time steps has enabled the agents to perform distributed updates without relying on information from agents on unobserved hypotheses in $\Theta \backslash \Theta_i$. Our contribution lies in proving that the normalization free updates lead to almost sure asymptotic convergence as shown in Section VI.

As it can be observed, agent i depends on neighbor estimates $\mu_{j,t}(\boldsymbol{\theta}), \forall j \in \mathcal{V}(\boldsymbol{\theta}), \forall \boldsymbol{\theta} \in \Theta_i$. The scalar term $A(\boldsymbol{\theta})_{ij}$ describes the weight assigned by agent i to the belief received from agent j at hypothesis $\boldsymbol{\theta}$. Each agent only computes $p_{i,T}(\boldsymbol{\theta})$ over the set Θ_i and depends on other agents for providing $p_{i,T}(\boldsymbol{\theta})$ for $\boldsymbol{\theta} \in \Theta \backslash \Theta_i$. At the final time step T, other agent estimates are used to obtain the probability values at hypothesis $\boldsymbol{\theta} \in \Theta \backslash \Theta_i$ for computing normalization factor $Z_{i,T}$. The update rules can also be expressed as a logarithm of the agent estimates at each discrete hypothesis $\boldsymbol{\theta}_v$, given as $q_{i,t}(\boldsymbol{\theta}_v) = \log(\mu_{i,t}(\boldsymbol{\theta}_v)); \mathrm{qz}_{i,t}(\boldsymbol{\theta}_v) = \log(\mathrm{pz}_i(\boldsymbol{z}_{i,t}|\boldsymbol{\theta}_v))$.

Algorithm 1: Partial likelihood estimation algorithm

Input: observations $\{\boldsymbol{z}_{i,t}\}_{t=1}^{T}$, hypothesis set Θ_i, prior hypothesis likelihoods $p_{i,0}(\boldsymbol{\theta})$ and communication matrices $\boldsymbol{A}(\boldsymbol{\theta})$ for all $\boldsymbol{\theta} \in \Theta_i$

Output: posterior probability $p_{i,T}$ for all $\boldsymbol{\theta} \in \Theta_i$

$\mu_{i,0}(\boldsymbol{\theta}) \leftarrow p_{i,0}(\boldsymbol{\theta}),\ \forall(\boldsymbol{\theta}) \in \Theta_i$

for $t \in \{1, \ldots, T-1\}$ **do**

  for $\boldsymbol{\theta} \in \Theta_i$ **do**

    $\mu_{i,t+1}(\boldsymbol{\theta}) = \prod_{j \in \mathcal{N}_i} \mu_{j,t}(\boldsymbol{\theta})^{A(\boldsymbol{\theta})_{ij}}\ \mathrm{pz}_i(\boldsymbol{z}_{i,t}|\boldsymbol{\theta})$

  end

end

$Z_{i,T+1} = \sum_{\boldsymbol{\theta} \in \Theta_i} \mu_{i,T+1}(\boldsymbol{\theta}) + \sum_{\boldsymbol{\theta} \in \Theta \setminus \Theta_i} \mu_{j,T+1}(\boldsymbol{\theta})$

$p_{i,T}(\boldsymbol{\theta}) = \mu_{i,T}(\boldsymbol{\theta}) / Z_{i,t+1}\ \forall \boldsymbol{\theta} \in \Theta_i$

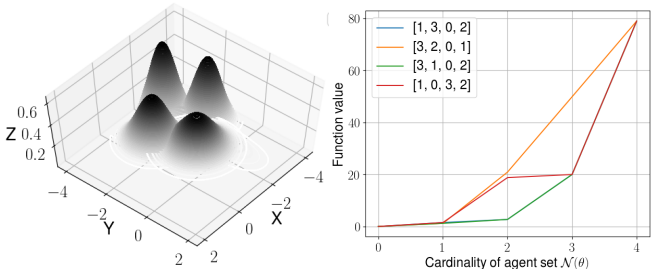

Fig. 2. True observation models $\mathrm{pz}_i(\boldsymbol{z}_{i,t}|\boldsymbol{\theta}) = N(\mathbf{x}_i - \boldsymbol{\theta}, ||x_i - \boldsymbol{\theta}||I_2)$. for 4 agents (left) and diversity function values $F(\mathcal{G})$ with varying number of set elements (right).

This allows for a more efficient numerical implementation. Define $q_{i,t+1}(\boldsymbol{\theta}_v) = \sum_{j=1}^{n} [A(\boldsymbol{\theta}_v)]_{ij} q_{j,t}(\boldsymbol{\theta}_v) + \mathrm{qz}_{i,t}(\boldsymbol{\theta}_v)$. This logarithm representation will be used extensively in the analysis section.

V. Assignment and Estimation Simulations

For illustrative purposes, an assignment simulation is carried out for a group of four agents communicating over a connected graph as shown in Fig. 3. The agent positions are $(-0.375, -1.125), (1.125, -0.375), (0.375, 1.125)$ and $(-1.125, 0.375)$. The circles represent nine hypotheses in a grid of $[-2, 2]^2$. The observation models are dependent on the agent positions, $\mathbf{x}_i \in \mathbb{R}^2$, and hypothesis locations, $\boldsymbol{\theta}_v \in \mathbb{R}^2$, and given as $\boldsymbol{z}_{i,t} \sim N(\mathbf{x}_i - \boldsymbol{\theta}_v, ||x_i - \boldsymbol{\theta}_v||I_2)$. The source representing the true value of joint parameter is placed at $\boldsymbol{\theta}^* = (2, 2)$. We plot the true observation models $f_i(\boldsymbol{z})$ of the form of 2D Gaussian distributions based on source $\boldsymbol{\theta}^*$ as $f_i^*(\boldsymbol{z}) \sim N(\mathbf{x}_i - \boldsymbol{\theta}^*, ||x_i - \boldsymbol{\theta}^*||I_2)$ in Fig. (2).

The cardinality constraint limits hypotheses tracked by each agent to 6 and the coverage constraint enforces that each hypothesis is observed by at least one agent. Considering the cycle formed by nodes $\{0, 1, 2\}$, we need to solve the complete integer optimization in Eqn. (7-9). Based on the number of nodes ($n = 4$), edges ($l = 4$) and hypotheses ($m = 9$), we optimize over $nm + lm + 2lm = 144$ scalar variables for this example. There are $n + 2m + 3ml = 130$ linear constraints and $ml = 36$ quadratic constraints. We optimized via Gurobi solver [23] using the barrier method and simplex method in conjunction with linear approximations.

The optimization yields the $\boldsymbol{y}^*, \boldsymbol{b}^*, \boldsymbol{f}^*$; out of which optimal agent assignment $\boldsymbol{y}^*$ is plotted via agent markers in Fig. (3) at each hypothesis. The assignment provides coverage as each hypothesis is tracked by at least one agent. Since the objective functions are defined in terms of the observation models $\mathrm{pz}_i(\cdot|\boldsymbol{\theta})$ depending on the distance between agents and hypotheses, the effect of diversity maximization leads to more agents observing farther hypotheses. For instance, the sensing agents $\{0, 2\}$ track hypotheses set $\{0, 1, 2\}$ and $\{6, 7, 8\}$ which are distant from the agents. The connectivity requirement is satisfied as every hypothesis is assigned an induced subgraph of agents. For instance, hypothesis 7 is observed by agents $\{0, 2\}$ connected by an edge. Since the diversity criterion is being maximized, the cardinality constraint is tight for each agent enabling tracking of exactly six hypotheses.

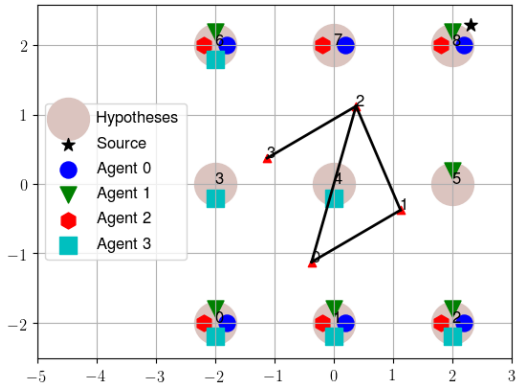

Fig. 3. Hypothesis assignment for 4 agents using distance-based observation models.

For the distributed estimation simulations, we pool agent-neighbor opinions via an assigned doubly stochastic matrix $\boldsymbol{A}(\theta)$ with positive diagonal elements at each of the hypothesis $\boldsymbol{\theta} \in \Theta$. In the simulation, we can observe the consensus among probability estimates of all agents at every hypothesis. The common probability values are maximized at the true hypothesis at $\boldsymbol{\theta} = (2, 2) \in \Theta^*$. Whereas at other hypotheses, the probabilities converge to zero. This illustrates our claim that even though agents observe a subset of up to 6 hypotheses as shown in Fig. 3, all the agent probability estimates $p_{i,t}(\boldsymbol{\theta})$ still converge to the correct value $p_{\infty}(\boldsymbol{\theta})$ at all hypotheses $\boldsymbol{\theta}$. The estimates reached the true consensus value without any intermediate normalization steps.

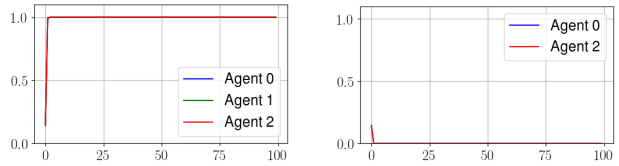

Fig. 4. Agent probability estimates $p_{i,t}(\boldsymbol{\theta})$ at all hypotheses upon running the algorithm for 100 time steps at true hypothesis $\boldsymbol{\theta} = [2, 2]$ (left) and at false hypothesis $\boldsymbol{\theta} = [0, 2]$ (right).

We present an application of this network wide estimation on a 20-node network looking for true hypothesis in a set of 900 hypotheses as shown in Fig. 5. We observe in the Fig. 6 (b) and (c) that the number of observed hypotheses by an agent can be diminished without any perceived difference in rate of convergence. In other simulations, it was observed

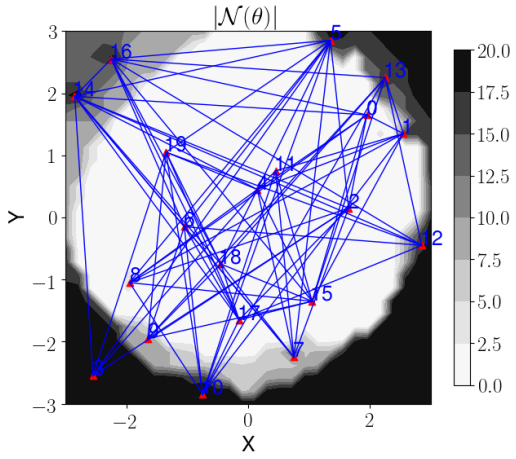

Fig. 5. Contour color showing number of agents observing each hypothesis location and the superimposed robot network.

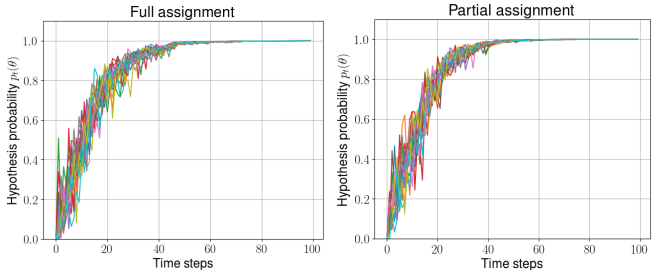

Fig. 6. Likelihood convergence at the true hypothesis when each agent observes all hypotheses (left) and only up to 300/900 hypotheses (right).

that the rate of convergence is not significantly affected on lowering the limit m_i on sets Θ_i as long as coverage criterion is satisfied. This validates our approach towards lowering the hypotheses tracked by each agent. For the 20-node network, the partial likelihood update with 300/900 tracked hypotheses results in same estimates with 1/3rd storage at each node and fewer than 1/3rd communication events in the network. We were able to deal with up to 3e5 variables on a machine with 16 GB RAM before running into memory overflow. These large scale simulations completed with both simplex and cutting plane methods in under 2 minutes.

VI. Proof of Asymptotic Convergence

We provide a proof of the almost-sure asymptotic convergence of the proposed partial likelihood estimation algorithm (Alg. 1) under the assumptions mentioned here:

Assumption 1 (Static graph)**.** The undirected graph $\mathcal{G}$ describing the agent communication is static and time-invariant.

Assumption 2 (Coverage)**.** Each hypothesis $\boldsymbol{\theta} \in \Theta$ is observed by at least one agent, i.e., $|\mathcal{V}(\boldsymbol{\theta})| \geq 1$.

Assumption 3 (Non-zero initial probabilities)**.** Every agent i has an initial likelihood $p_{i,0}(\boldsymbol{\theta}) > 0$, $\forall \boldsymbol{\theta} \in \Theta_i$.

Assumption 4 (Sensor data reception)**.** Assume that if the true data generating pdf for agent i, $f_i^*(\boldsymbol{z}) > 0$ for some $z \in \mathbb{R}^{d_z}$, then $1 \geq \bar{\alpha} \geq \mathrm{pz}_i(z|\boldsymbol{\theta}) \geq \underline{\alpha} > 0$, for all $\boldsymbol{\theta} \in \Theta_i$ and some constants $\bar{\alpha}$, $\underline{\alpha}$. Note that $\bar{\alpha}$ exists for any pdf.

As shown in Section III-C, we can assign a doubly-stochastic matrix $A(\boldsymbol{\theta}_v)$ to any hypothesis $\boldsymbol{\theta}_v \in \Theta$. Since every agent receives data from some $\boldsymbol{\theta} \in \Theta^*$, the identity of the agent tracking the likelihood of a true hypothesis is irrelevant. Let the number of agents observing hypothesis $\boldsymbol{\theta}_v$ be $n_v = |\mathcal{V}(\boldsymbol{\theta}_v)|$. Define also the vector of log-likelihoods of all observations received by the agents at time t given $\boldsymbol{\theta}_v$:

$$\mathrm{qz}_t(\boldsymbol{\theta}_v) := \begin{bmatrix} \log \mathrm{pz}_1(\boldsymbol{z}_{1,t}|\boldsymbol{\theta}_v) \\ \vdots \\ \log \mathrm{pz}_{n_v}(\boldsymbol{z}_{n_v,t}|\boldsymbol{\theta}_v) \end{bmatrix}.$$

Similarly, define $q_0(\boldsymbol{\theta}_v) := \log(p_0(\boldsymbol{\theta}_v))$ as the vector of initial likelihood across all agents assigned to hypothesis $\boldsymbol{\theta}_v$, which is well defined due to Assumption 3.

Lemma 1. *Under Assumptions 1 and 4, the asymptotic likelihood estimates of all agents in set $\mathcal{V}(\boldsymbol{\theta}_v)$ observing hypothesis $\boldsymbol{\theta}_v$ are equal, i.e., for some $c \in \mathbb{R}_{>0}$:*

$$\lim_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} \boldsymbol{A}(\boldsymbol{\theta}_v)^{T-t}\, \mathrm{qz}_t(\boldsymbol{\theta}_v) = c\mathbf{1}_{n_v} \qquad (12)$$

Proof. Using the ergodicity property of doubly-stochastic matrices, $\lim_{T\to\infty} \boldsymbol{A}(\boldsymbol{\theta}_v)^T = \frac{1}{n_v}\mathbf{1}_{n_v}\mathbf{1}_{n_v}^\top$, and Assumption 4 that $\log(\bar{\alpha})\mathbf{1}_{n_v} \geq \mathrm{qz}_t(\boldsymbol{\theta}_v) \geq \log(\underline{\alpha})\mathbf{1}_{n_v}$, $\forall\, t \geq 0$, we have:

$$\lim_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} \left(\boldsymbol{A}(\boldsymbol{\theta}_v)^t - \frac{1}{n_v}\mathbf{1}_{n_v}\mathbf{1}_{n_v}^\top \right) \mathrm{qz}_{T-t}(\boldsymbol{\theta}_v) = 0. \quad (13)$$

Upon reversing the time index in the summation in (12),

$$\frac{1}{T} \sum_{t=0}^{T-1} \boldsymbol{A}(\boldsymbol{\theta}_v)^{T-t}\, \mathrm{qz}_t(\boldsymbol{\theta}_v) = \frac{1}{T} \sum_{t=1}^{T} \boldsymbol{A}(\boldsymbol{\theta}_v)^{t}\, \mathrm{qz}_{T-t}(\boldsymbol{\theta}_v)$$

Now, let us take the limit $T \to \infty$ and use the ergodicity property of matrix $\boldsymbol{A}$ on the term in Eqn. (12) to obtain

$$\frac{1}{n_v}\mathbf{1}_{n_v} \left[\lim_{T\to\infty} \frac{1}{T} \sum_{t=0}^{T-1} \left(\sum_{j\in\mathcal{V}(\boldsymbol{\theta}_v)} \log(\mathrm{pz}_j(\boldsymbol{z}_{j,t}|\boldsymbol{\theta}_v)) \right) \right].$$

If the limit of the sum of observation log likelihoods above exists, then the asymptotic estimates of every agent $i \in \mathcal{V}(\boldsymbol{\theta}_v)$ would converge to the same value. We will use Kolmogorov's strong law of large numbers to show the existence of this limit. The strong law states that $\frac{1}{T}\sum_{t=1}^{T} X_t \to \frac{1}{T}\sum_{t=1}^{T} \mathbb{E}[X_t]$ almost surely for a sequence of independent random variables $\{X_t\}$ with expectations $\mathbb{E}[X_t] < \infty$ and variances $\sum_{t=1}^{\infty} Var(X_t)/t^2 < \infty$. Letting $X_t := \sum_{j\in\mathcal{V}(\boldsymbol{\theta}_v)} \log(\mathrm{pz}_j(\boldsymbol{z}_{j,t}|\boldsymbol{\theta}_v))$, we can verify that both conditions required by the strong law are satisfied due to Assumption 4. Considering the definition of expectation as sum over temporal observations, we get,

$$\frac{1}{n_v}\mathbf{1}_{n_v} \left(\sum_{j\in\mathcal{V}(\boldsymbol{\theta}_v)} \mathbb{E}[\log(\mathrm{pz}_j(z|\boldsymbol{\theta}_v))] \right) \qquad (14).$$

This shows that all agents arrive at the same asymptotic estimate of the unnormalized likelihood of hypothesis $\boldsymbol{\theta}_v$. As a result, the normalization factors $Z_{i,t}$ in (11) also asymptotically converge to the same value across all agents. □

To proceed, we analyze the constant c in (12) further. The expectations in (14) can be expressed in terms of the entropy and the KL-divergence with respect to the true observation

model $f_i(\cdot)$ of the agent observation models $\mathrm{pz}_i(\cdot|\boldsymbol{\theta}_v)$:

$$\mathbb{E}\left[\log(\mathrm{pz}_i(z|\boldsymbol{\theta}_v))\right] = \int f_i(\boldsymbol{z}) \log\left(\mathrm{pz}_i(z|\boldsymbol{\theta}_v)\frac{f_i(\boldsymbol{z})}{f_i(\boldsymbol{z})}\right) dz$$
$$= -\mathrm{D}_{\mathrm{KL}}(f_i(\cdot)||\,\mathrm{pz}_i(\cdot|\boldsymbol{\theta}_v)) - \mathrm{H}(f_i(\cdot)). \quad (15)$$

We use the notation $[\,]_i$ to denote the entry in a vector corresponding to agent i. Based on Lemma 1, eq. (15), and $\lim_{t\to\infty} \boldsymbol{A}(\boldsymbol{\theta}_v)^T = \frac{1}{n_v}\mathbf{1}_{n_v}\mathbf{1}_{n_v}^\top$, thus averaging the prior likelihood as $T \to \infty$, we have:

$$\lim_{T\to\infty} \frac{1}{T}\left[\boldsymbol{A}(\boldsymbol{\theta}_v)^T \log(p_0(\boldsymbol{\theta}_v)) + \sum_{t=1}^{T} \boldsymbol{A}(\boldsymbol{\theta}_v)^{T-t}\,\mathrm{qz}_t(\boldsymbol{\theta}_v)\right]_i$$
$$= \lim_{T\to\infty} \frac{1}{T}\left[\sum_{t=1}^{T} \boldsymbol{A}(\boldsymbol{\theta}_v)^{T-t}\,\mathrm{qz}_t(\boldsymbol{\theta}_v)\right]_i, \quad (16)$$
$$= \frac{1}{n_v}\left(\sum_{j\in\mathcal{V}(\boldsymbol{\theta}_v)} -\mathrm{D}_{\mathrm{KL}}(f_j(\cdot)||\,\mathrm{pz}_j(\cdot|\boldsymbol{\theta}_v)) - \mathrm{H}(f_j(\cdot))\right).$$

Finally, we prove convergence of the estimates maintained by agent i at hypotheses in Θ_i to a density function with non-zero mass only over Θ^*. We rely on the difference in the convergence rates of the likelihood $p_{i,t}(\cdot)$ evaluated at a true hypothesis $\boldsymbol{\theta}_w \in \Theta^*$ versus a false hypothesis $\boldsymbol{\theta}_v \notin \Theta^*$. Showing that the probability ratio, $p_{i,t}(\boldsymbol{\theta}_v)/p_{i,t}(\boldsymbol{\theta}_w)$, converges to zero is enough to guarantee that the probability mass at an incorrect hypothesis, $\boldsymbol{\theta}_v$, is asymptotically zero. Let the log probability ratio of $\boldsymbol{\theta}_v$ and $\boldsymbol{\theta}_w$ for agent i be:

$$\phi_{i,T+1}(\boldsymbol{\theta}_v,\boldsymbol{\theta}_w) = \log\frac{\mu_{i,T+1}(\boldsymbol{\theta}_v)/Z_{i,T+1}}{\mu_{i,T+1}(\boldsymbol{\theta}_w)/Z_{i,T+1}},$$
$$= \log\frac{\prod_{j\in\mathcal{V}(\boldsymbol{\theta}_v)} p_{j,T}(\boldsymbol{\theta}_v)^{\boldsymbol{A}(\boldsymbol{\theta}_v)_{ij}}\,\mathrm{pz}_i(\boldsymbol{z}_{i,T+1}|\boldsymbol{\theta}_v)}{\prod_{j\in\mathcal{V}(\boldsymbol{\theta}_w)} p_{j,T}(\boldsymbol{\theta}_w)^{\boldsymbol{A}(\boldsymbol{\theta}_w)_{ij}}\,\mathrm{pz}_i(\boldsymbol{z}_{i,T+1}|\boldsymbol{\theta}_w)},$$
$$= \left[\sum_{t=1}^{T} \boldsymbol{A}(\boldsymbol{\theta}_v)^{T-t}\,\mathrm{qz}_t(\boldsymbol{\theta}_v) + \boldsymbol{A}(\boldsymbol{\theta}_v)^t q_0(\boldsymbol{\theta}_v)\right]_i \quad (17)$$
$$- \left[\sum_{t=1}^{T} \boldsymbol{A}(\boldsymbol{\theta}_w)^{T-t}\,\mathrm{qz}_t(\boldsymbol{\theta}_w) + \boldsymbol{A}(\boldsymbol{\theta}_w)^t q_0(\boldsymbol{\theta}_w)\right]_i.$$

Distributed inference algorithms based on Bayesian updates with normalization [24] optimize an objective function:

$$C(\boldsymbol{\theta}_v) = \frac{1}{n_v}\sum_{j\in\mathcal{V}(\boldsymbol{\theta}_v)} \left(\mathrm{D}_{\mathrm{KL}}(f_j(\cdot)||\,\mathrm{pz}_j(\cdot|\boldsymbol{\theta}_v)) + \mathrm{H}(f_j(\cdot))\right)$$

Following Eqn. (16), the individual terms exist in the time-averaged asymptotic value, $\lim_{T\to\infty}\frac{1}{T}\phi_{i,T+1}(\boldsymbol{\theta}_v,\boldsymbol{\theta}_w)$,

$$\lim_{T\to\infty}\frac{1}{T}\phi_{i,T+1}(\boldsymbol{\theta}_v,\boldsymbol{\theta}_w) = C(\boldsymbol{\theta}_w) - C(\boldsymbol{\theta}_v) \quad (18)$$

This is related to (16) since for $\boldsymbol{\theta}_v \notin \Theta^*$ and $\boldsymbol{\theta}_w \in \Theta^*$,

$$\lim_{T\to\infty}\frac{1}{T}\phi_{i,T+1}(\boldsymbol{\theta}_v,\boldsymbol{\theta}_w) = C^* - C(\boldsymbol{\theta}_v) < 0. \quad (19)$$

Therefore, the value of $\phi_{i,t}(\boldsymbol{\theta}_v,\boldsymbol{\theta}_w) \to -\infty$ almost surely. Also since $p_{i,t}(\boldsymbol{\theta}_v) \leq \exp(\phi_{i,t}(\boldsymbol{\theta}_v,\boldsymbol{\theta}_w))$, $\forall i \in \mathcal{V}(\theta_v) \cap \mathcal{V}(\theta_w)$, we have $p_{i,t}(\boldsymbol{\theta}_v) \to 0$. Therefore, the likelihood of an incorrect hypothesis is asymptotically almost surely zero.

VII. Conclusion

In this paper, we have proposed a distributed inference algorithm to allow sensor networks to learn source distribution while maintaining partial observation likelihoods. This leads to significant savings in the number of message exchanged across neighbors. The devised algorithms has proven convergence guarantees in the absence of normalization factors at each update step, thus allowing a significant speed up of the algorithm. As the distributed estimation algorithm depends on hypothesis agent matching, we have also devised a novel mixed integer programming formulation for assigning connected subgraphs of agents to each hypothesis. The results of asymptotic convergence of the algorithm for fixed graphs can also be extended to the cases of asynchronous infinite-often communications and time-varying graphs with global connectivity over certain fixed time steps.

References

[1] F. Büsching, S. Schildt, and L. Wolf. Droidcluster: Towards smartphone cluster computing–the streets are paved with potential computer clusters. In *International Conference on Distributed Computing Systems Workshops*, pages 114–117. IEEE, 2012.

[2] N. Atanasov, R. Tron, V.M. Preciado, and G.J. Pappas. Joint estimation and localization in sensor networks. In *IEEE Conference on Decision and Control*, pages 6875–6882. IEEE, 2014.

[3] G. Zyskind, O. Nathan, and A. Pentland. Enigma: Decentralized computation platform with guaranteed privacy. *arXiv:1506.03471*, 2015.

[4] R.J. Aumann. Agreeing to disagree. *The Annals of Statistics*, pages 1236–1239, 1976.

[5] J. Tsitsiklis, D. Bertsekas, and M. Athans. Distributed asynchronous deterministic and stochastic gradient optimization algorithms. *IEEE Transactions on Automatic Control*, 31(9):803–812, 1986.

[6] M. Alanyali, S. Venkatesh, O. Savas, and S. Aeron. Distributed Bayesian hypothesis testing in sensor networks. In *American Control Conference (ACC)*, pages 5369–5374 vol.6, 2004.

[7] S. Kar, J. M. F. Moura, and K. Ramanan. Distributed parameter estimation in sensor networks: Nonlinear observation models and imperfect communication. *IEEE Transactions on Information Theory*, 58(6):3575–3605, June 2012.

[8] A. Jadbabaie, P. Molavi, A. Sandroni, and A. Tahbaz-Salehi. Non-bayesian social learning. *Games and Economic Behavior*, 76(1):210–225, 2012.

[9] A. Nedić, A. Olshevsky, and C. A. Uribe. Nonasymptotic convergence rates for cooperative learning over time-varying directed graphs. In *American Control Conference (ACC)*, pages 5884–5889, 2015.

[10] A. Nedić, A. Olshevsky, and C.A. Uribe. Fast convergence rates for distributed non-bayesian learning. *IEEE Transactions on Automatic Control*, 62(11):5538–5553, 2017.

[11] H.W. Kuhn. The hungarian method for the assignment problem. *Naval research logistics quarterly*, 2(1-2):83–97, 1955.

[12] D.W. Pentico. Assignment problems: A golden anniversary survey. *European Journal of Operational Research*, 176(2):774–793, 2007.

[13] P. Biswas, T. Lian, T. Wang, and Y. Ye. Semidefinite programming based algorithms for sensor network localization. *ACM Transactions on Sensor Networks (TOSN)*, 2(2):188–220, 2006.

[14] E. Álvarez-Miranda, I. Ljubić, and P. Mutzel. The maximum weight connected subgraph problem. In *Facets of Combinatorial Optimization*, pages 245–270. Springer, 2013.

[15] A.A. Loboda, M.N. Artyomov, and A.A. Sergushichev. Solving generalized maximum-weight connected subgraph problem for network enrichment analysis. In *International Workshop on Algorithms in Bioinformatics*, pages 210–221. Springer, 2016.

[16] B. Touri. *Product of random stochastic matrices and distributed averaging*. Springer Science & Business Media, 2012.

[17] L. Lovász. *Submodular functions and convexity*, pages 235–257. Springer Berlin Heidelberg, Berlin, Heidelberg, 1983.

[18] S.T. McCormick. Submodular function minimization. *Handbooks in Operations Research and Management Science*, 12:321–391, 2005.

[19] D.S. Johnson. The np-completeness column: an ongoing guide. *Journal of Algorithms*, 6(3):434 – 451, 1985.

[20] Y. Wang, A. Buchanan, and S. Butenko. On imposing connectivity constraints in integer programs. *Mathematical Programming*, 166(1-2):241–271, 2017.

[21] N. Cohen. Several graph problems and their linear program formulations. Working paper or preprint, January 2019.

[22] R. Sinkhorn and P. Knopp. Concerning nonnegative matrices and doubly stochastic matrices. *Pacific Journal of Mathematics*, 21(2):343–348, 1967.

[23] LLC Gurobi Optimization. Gurobi optimizer reference manual, 2018.

[24] S.G. Walker. Bayesian inference via a minimization rule. *Sankhyā: The Indian Journal of Statistics*, pages 542–553, 2006.