Distributed estimation algorithms on optimally assigned hypotheses

Parth Paritosh, Nikolay Atanasov and Sonia Martinez

***Abstract*— We design a cooperative, decentralized inference algorithm allowing sensor networks to learn a probability density over a discrete set of hypotheses representing a joint parameter that best explains their combined observations. 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 under restricted 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. Second, our algorithm does not require step-wise renormalization across agents, while still guaranteeing consensus and convergence of sensor estimates. Third, we also address agent-hypothesis assignment by formulating it as an integer programming problem, that matches agent sub-networks to hypotheses based on a diversity criterion for estimation quality.**

I. Introduction

Distributed estimation algorithms have long been studied from a 2-agent agreement [1] to studies on network topology among multiple agents [6]. A major improvement appeared in the form of non-Bayesian updates [2], 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 [4]. Even though recent approaches can deal with network-level updates, they require maintaining and communicating each agents pdf over the entire set of hypotheses. Drawing inspiration from the idea of distributed computation, it is of interest and practically useful to also consider a distributed storage scheme, in which agents only maintain and exchange a partial pdf over the parameter space. This naturally raises the question of defining hypothesis assignments for agents, which is a problem akin to Sensor selection. Load assignment for sensor selection at is usually studied as an assignment problem [5], mostly tackled via integer-programming solutions. With increasing performance requirements, there is a need to develop new techniques and representations. We present our novel proposals for agent-hypothesis assignment and partial space distributed assignment problems.

II. Problem Formulation

We consider a set of sensors $\mathcal{N} = \{1, \dots, n\}$ whose communications are modeled via an undirected graph $\mathcal{G} = (\mathcal{N}, \mathcal{E})$, 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. The observation model is specified for sensor i by a pdf $\mathrm{pz}_i(\boldsymbol{z}|\boldsymbol{\theta})$ defined on discrete domain Θ_i. The network aims to find the true value of a joint parameter $\boldsymbol{\theta} \in \mathbb{R}^{d_\theta}$ from the set $\Theta \equiv \cup_{i\in\mathcal{N}}\Theta_i$, which may represent the location of a data-generating source. The 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 $\boldsymbol{\theta} \in \Theta_i$.

Sub-network assignment for hypotheses. We can now formalize the implications of the restriction imposed by the set Θ_i. This set limits the computational load of agent i by restricting its likelihood over $|\Theta_i| \leq m_i$ hypotheses. The coverage of entire hypothesis space is ensured with the condition $\cup_{i\in\mathcal{N}}\Theta_i = \Theta$ implying that every hypothesis is tracked by at least one agent in the network. 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 component of $\mathcal{G}$, we define subgraph $\mathcal{G}_{\theta_v} \equiv \mathcal{G}_v$. The connectedness is desirable to achieve consensus on estimates at each hypothesis. Based on these constraints, we define the diversity criterion using distance between observation models as $F(\mathcal{G}_{\boldsymbol{\theta}}) = \sum_{\substack{i,j\in\mathcal{V}(\boldsymbol{\theta})\\(i,j)\in\mathcal{E}}} \mathrm{D}_{\mathrm{KL}}(\mathrm{pz}_i(\cdot|\boldsymbol{\theta}), \mathrm{pz}_j(\cdot|\boldsymbol{\theta})) + \sum_{i\in\mathcal{V}(\boldsymbol{\theta})} \mathrm{H}(\mathrm{pz}_i(\cdot|\boldsymbol{\theta}))$. Thus, our first research question is addressed by means of an optimization problem over sets of subgraphs $\mathcal{G}_{\boldsymbol{\theta}} \subseteq \mathcal{G}, \forall \boldsymbol{\theta} \in \Theta$,

$$\max_{\{\mathcal{G}_{\boldsymbol{\theta}}\}_{\boldsymbol{\theta}\in\Theta}} \sum_{\boldsymbol{\theta}\in\Theta} F(\mathcal{G}_{\boldsymbol{\theta}}), \tag{1}$$

$$\mathcal{G}_{\boldsymbol{\theta}} \text{ is a connected induced subgraph of } \mathcal{N},\ \forall \boldsymbol{\theta} \in \Theta, \tag{2}$$

$$0 < |\Theta_i| \leq m_i\ \forall i \in \mathcal{N}, \tag{3}$$

$$\cup_i \Theta_i = \Theta. \tag{4}$$

Distributed inference on limited hypothesis sets. In this problem, each agent is tasked with finding the values of a probability mass function $p_{i,T}(\boldsymbol{\theta})$ for the hypotheses $\boldsymbol{\theta} \in \Theta_i$ at time T using only neighbor estimates and received observations $\boldsymbol{z}_{i,1:T}$. For the values $\boldsymbol{\theta} \in \Theta \setminus \Theta_i$ at time T, the network will collectively assign a complementary mass value. 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.

III. Agent-hypothesis matching

The closest problem to the optimization problem in Eqn. (1-4) is the reference on generalized maximum-weight connected subgraph problems [3]. Although the solution finds connected sensor sub-network with weights on nodes and edges, it isn't

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

scaled for several subgraphs coupled with cardinality constraints on selected agents. Therefore, we present a new formulation in terms of binary variables representing nodes and edges in the optimal subgraph. 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 $\mathbf{b}_v$. Exploiting the separability of the objective function on each $\boldsymbol{\theta}_v$, we define the linear objective in terms of the new node and edge binary variables.

$$\sum_{v=1}^{m} \Big[\max_{\mathbf{y}_v, \mathbf{b}_v} \sum_{(i,j)\in\mathcal{E}} b_{ij,v} \, \mathrm{D_{KL}}(\mathrm{pz}_i(z|\boldsymbol{\theta}_v), \mathrm{pz}_j(z|\boldsymbol{\theta}_v)) + \sum_{i=1}^{n} y_{iv} \, \mathrm{H}(\mathrm{pz}_i(z|\boldsymbol{\theta}_v)) \Big] \tag{5}$$

$$\sum_{v=1}^{m} y_{iv} \le m_i, \quad \forall i \in \{1,\ldots,n\}, \tag{Cardinality}$$

$$\sum_{i} y_{iv} \ge 1, \quad \forall v \in \{1,\ldots,m\}. \tag{Coverage}$$

In addition, the first set of graph connectivity constraints can be expressed by adding constraints on the edges for each hypothesis $v \in \{1,\ldots,m\}$ as in Eqn. (6).

$$\sum_{ij\in\mathcal{E}} b_{ij,v} = \sum_{i=1}^{n} y_{iv} - 1, \; b_{ij,v} \le y_{iv}, y_{jv}, \; \forall ij \in \mathcal{E}. \tag{6}$$

We can further note that nodes i,j defining a selected edge (i,j) with $b_{ij,v} = 1$ are automatically selected, i.e. $y_{iv}, y_{jv} = 1$. This is seen with arguments relating number of nodes and edges in the Eqn. (6) suffices if the original graph is a tree. Otherwise, one can rely on introducing average node degree constraints. The average node degree of a network is defined as the average over all node degrees in the network. Therefore, the maximum average degree of a connected tree of k nodes is $2 - 2/k$. This idea is harnessed by introducing flow variables $f^i_{ij,v}, f^j_{ij,v} \in \{0,1\}$ on each edge $(i,j) \in \mathcal{E}$ with $f^i_{ij,v} + f^j_{ij,v} = 2$, and another quadratic constraint on the average degree of the optimal subgraph $\sum_{j\in\mathcal{N}_i} f^j_{ij,v} \le 2 - \frac{2}{\sum_{i=1}^n y_{iv}}$ to guarantee connectivity in selected sets at hypothesis $\theta_v \in \{1,\ldots,m\}$.

IV. Distributed consensus on partial hypothesis

In this section, we propose a network-wide inference algorithm that can be implemented with partial observation likelihoods for each agent without any network-wide renormalization. As it can be observed in Algorithm 1, agent i depends on estimates $\mu_{j,t}(\boldsymbol{\theta}), \forall j \in \mathcal{V}(\boldsymbol{\theta}), \boldsymbol{\theta} \in \Theta_i$. 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 updates $p_{i,T}(\boldsymbol{\theta})$ over the set Θ_i and depends on other agents for obtaining $p_{i,T}(\boldsymbol{\theta})$ 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 $Z_{i,T}$.

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

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

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

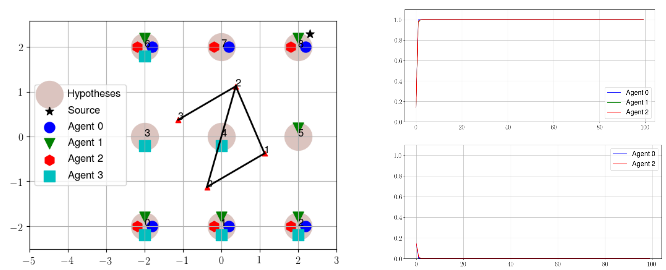

Fig. 1. (a) Hypothesis assignments for agents positioned at $(-0.375, -1.125)$, $(1.125, -0.375)$, $(0.375, 1.125)$ and $(-1.125, 0.375)$. The 9 hypotheses are equally spaced in a grid of $[-2,2]^2$. The observation models are given as $\mathrm{pz}_i(\boldsymbol{z}_{i,t}|\boldsymbol{\theta}) \sim N(\mathbf{x}_i - \boldsymbol{\theta}_v, \frac{1}{||x_i - \boldsymbol{\theta}_v||} I_2)$. Consistent agent probability estimates $p_{i,t}(\boldsymbol{\theta})$ at all hypotheses upon running the algorithm for 100 time steps at (b) true hypothesis $\boldsymbol{\theta} = [2,2]$ and at (c) false hypothesis $\boldsymbol{\theta} = [0,2]$.

V. Conclusion

In this work, we have proposed a distributed inference algorithm that allows each sensor to work with partial observation likelihoods, leading to significant savings in the number of hypotheses stored at each sensor and the messages exchanged among neighbors. The devised algorithms has proven convergence guarantees in the absence of normalization factors at each update step. As the distributed estimation algorithm depends on hypothesis agent matching, we have also devised a novel integer programming formulation for assigning connected subgraphs of agents to each hypothesis.

References

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

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

[3] 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.

[4] 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.

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

[6] 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.