Robust Asynchronous and Network-Independent Cooperative Learning

Eduardo Mojica-Nava†, David Yanguas-Rojas†, Cesar A. Uribe‡

†Universidad Nacional de Colombia, ‡LIDS-MIT

Introduction

Distributed inference, and in particular non-Bayesian social learning [2], has gained increasing attention in recent years due to the numerous applications in machine learning, sensor networks, decentralized control, and distributed signal processing [3]. In this paper, we build upon recently available results in distributed optimization considering asynchrony, delays, and message losses [6, 5, 1], and introduce a cooperative distributed non-Bayesian learning algorithm with robust performance guarantees under such harsh communication network conditions. In particular, we extend the recent proposed Robust Asynchronous Push-Sum (RAPS) consensus algorithm [4] to the distributed learning setup.

The main contribution of this paper is threefold:

- We introduce a robust distributed non-Bayesian cooperative learning algorithm that considers asynchronous updates, communication delays, and unpredictable messages losses over a directed communication graph. A network of agents tries to jointly agree on a hypothesis that best describes a sequence of locally and asynchronously available observations.
- We show that our proposed learning dynamics guarantee that all agents in the network will have an asymptotic exponential decay of their beliefs on the wrong hypothesis, indicating that the beliefs of all agents will concentrate on the optimal hypotheses.
- We present numerical experiments that provide evidence of the proposed algorithm's performance on a number of network setups.

Cooperative Learning: Problem Statement

Consider a network of n agents on a set of nodes $V = \{1, 2, \dots, n\}$ observing realizations of a finite, stationary, independent, identically distributed random processes $\{X_k\}_{k\geq 1}$ where $X_k^i \sim P^i$ at each iteration time k with unknown distribution P^i. Additionally, all agents have a shared finite set of hypotheses $\Theta = \{\theta_1, \theta_2, \dots, \theta_m\}$, from which each agent $i \in V$ defines a local family of distributions $\mathcal{P}^i = \{P_\theta^i \mid \theta \in \Theta\}$. We will assume the technical condition that each element in family of distributions $\mathcal{P}^i$ is absolutely continuous with respect to P^i. We denote $\mathcal{N}_i^+$ and $\mathcal{N}_i^-$ as the set of out-neighbors and in-neighbors of an agent i. The objective of the network of agents is to agree on a parameter $\theta^* \in \Theta$ such that the joint distribution $\prod P_{\theta^*}^i$ is closest to $\prod P^i$. Formally, the group of agents tries to solve jointly

$$\min_{\theta\in\Theta} F(\theta) \triangleq \sum_{i\in V} D_{KL}(P^i \| P_\theta^i),$$

where $D_{KL}(P\|Q)$ is the Kullback-Leibler divergence between the distributions P and Q. Each agent only knows its local family of distributions $\mathcal{P}^i$, the true distribution of their local observations P^i is unknown yet accessible via local observations.

Objective: agents have to agree on a parameter $\theta^* \in \Theta$ such that the joint distribution $\prod P_{\theta^*}^i$ is closest to $\prod P^i$.

Network Model

We will generally denote the set of minimizers as Θ^*. The confidence each agent has on each of the hypotheses in Θ is represented by a *belief* vector, denoted as $\mu_i^\theta(k)$, which indicates the belief that an agent $i \in V$ has about a hypothesis $\theta \in \Theta$ at certain time instant k. A value of $\mu_i^\theta(k) = 1$ indicates certainty that the minimizer is θ, whereas $\mu_i^\theta(k) = 0$ indicates certainty that it is not. Agents cooperate by communicating their beliefs at each time instant. Such communication is mediated by a network, modeled as a graph $\mathcal{G} = \{V, E\}$.

We assume that the graph $\mathcal{G}$ is strongly connected and does not have self-loops, the delays on each link are bounded above by some $L_{del} \geq 1$, every agent wakes up and performs updates at least every $L_u \geq 1$ iterations, each link fails at most $L_f \geq 1$ consecutive times, and messages arrive in the order of time they were sent. **Assumption 1** in [4] Suppose:

(a) Graph $\mathcal{G}$ is strongly connected and does not have self-loops.

(b) The delays on each link are bounded above by some $L_{del} \geq 1$.

(c) Every agent wakes up and performs updates at least every $L_u \geq 1$ iterations.

(d) Each link fails at most $L_f \geq 1$ consecutive times.

(e) Messages arrive in the order of time they were sent. In other words, if message is sent from node i to node j at time k_1 and k_2, respectively, with effective delays d_1 and d_2, then $k_1 + d_1 < k_2 + d_2$.

Assumption 1 is, to the best of the author's knowledge, the weakest assumptions available in the literature for cooperative learning. It allows finite communication delays between agents and possible link failures, and asynchronous activation of the nodes. However, we impose the useful assumption that messages will arrive in the order they were sent.

Robust Asynchronous Push-Sum Distributed Non-Bayesian Learning

We state Algorithm 1, our proposed cooperative learning algorithm, and state our main results.

Algorithm 1 Robust Asynchronous Push-Sum Distributed Non-Bayesian Learning

Initialize: $y_i(0) = 1$, $\phi_i^y(0) = 0$, $\phi_{i,\theta}^\mu(0) = 1$, $\forall i \in V$, and $\rho_{ij}^y(0) = 0$, $\kappa_{ij}(0) = 0, \forall (i,j) \in \mathcal{E}$

Set initial beliefs as uniform for all agents.

for $k = 0, 1, 2, \dots,$ for every node i: **do**

  if Node i wakes up **then**

  1. Processing and broadcasting local information

    $\kappa_i \leftarrow k$, $\phi_i^y \leftarrow \phi_i^y + y_i/(d_i^+ + 1)$

    $\phi_{(i,\theta)}^\mu \leftarrow \phi_{(i,\theta)}^\mu \left(\mu_i^\theta\right)^{y_i/(d_i^+ + 1)}$

    Node i broadcasts $(\phi_i^y, \phi_{(j,\theta)}^\mu, \kappa_i)$ to $\mathcal{N}_i^+$.

  2. Processing received messages

    for $(\phi_j^y, \phi_{(i,\theta)}^\mu, \kappa_j')$ in the inbox **do**

      if $\kappa_j' > \kappa_{ij}$ **then**

        $\rho_{ij}^{*y} \leftarrow \phi_j^y$, $\rho_{ij|\theta}^{*\mu} \leftarrow \phi_{(j,\theta)}^\mu$, $\kappa_{ij} \leftarrow \kappa_j'$

      end if

    end for

  3. Updating beliefs and local information

    $\hat{y}_i \leftarrow \frac{y_i}{d_i^+ + 1} + \sum_{j\in\mathcal{N}_i^-} (\rho_{ij}^{*y} - \rho_{ij}^y)$

    $\mu_i^\theta \leftarrow \frac{1}{Z_i}\left(\left(\mu_i^\theta\right)^{\frac{y_i}{d_i^+ + 1}} \prod_{j\in\mathcal{N}_i^-} \left(\frac{\rho_{ij|\theta}^{*\mu}}{\rho_{ij|\theta}^\mu}\right) P_\theta^i(x_{k+1}^i)\right)^{\frac{1}{\hat{y}_i}}$

  Z_i is a normalization constant.

    $y_i \leftarrow \hat{y}_i$, $\rho_{ij}^y \leftarrow \rho_{ij}^{*y}$, $\rho_{ij|\theta}^\mu \leftarrow \rho_{ij|\theta}^{*\mu}$

  end if

end for

Fig. 1: Algorithm 1

Main Result

Theorem 1 (Main Result)**.** *Let assumptions about the connectivity of the communication network hold. Then, the output of Algorithm 1 has the following property:*

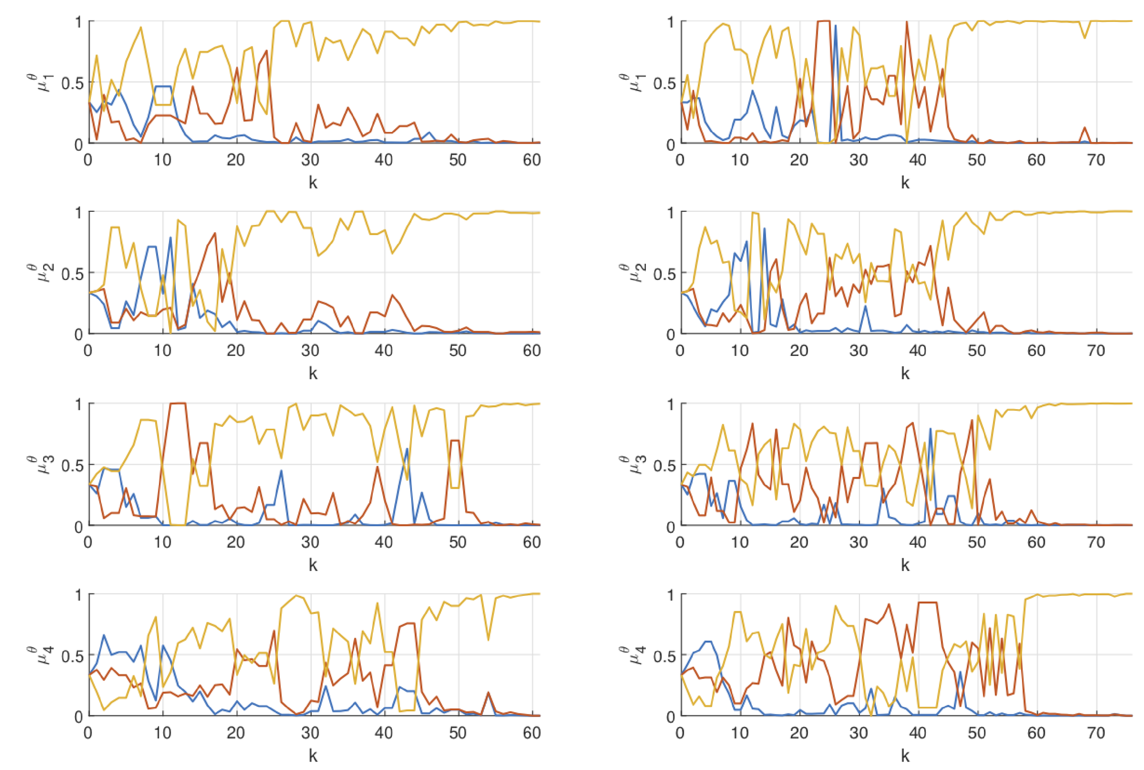
$$\lim_{k\to\infty} \frac{1}{k} \log \frac{\mu_{\theta_v}^i(k)}{\mu_{\theta_w}^i(k)} \leq -\frac{1}{n} \min_{\theta\notin\Theta^*} (F(\theta) - F(\theta^*))$$

almost surely for all $\theta_v \notin \Theta^*$, and $\theta_w \in \Theta^*$, and $i \in V$.

Remark: Note that the result in Theorem (Main Result) implies that $\lim_{k\to\infty} \mu_k^i(\theta) = 0$ almost surely for all $\theta \notin \Theta^*$.

Numerical Experiments

We present numerical experiments to illustrate the behavior of Algorithm 1 under the network model that allows asynchronous updates, delays, and link failures. We take a network with $n = 4$ agents connected through two different base topologies: a path (left) and a start (right), with waking up and failure probabilities of 0.9 and 0.2, respectively.



References

[1] Christoforos N Hadjicostis, Alejandro D Dominguez-Garcıa, and Themistokis Charalambous. "Distributed Averaging and Balancing in Network Systems". In: *Foundations and Trends in Systems and Control* 5.2-3 (2018), pp. 99–292.

[2] Ali Jadbabaie et al. "Non-Bayesian social learning". In: *Games and Economic Behavior* 76.1 (2012), pp. 210–225. ISSN: 0899-8256.

[3] Angelia Nedić, Alex Olshevsky, and César A Uribe. "A tutorial on distributed (non-Bayesian) learning: Problem, algorithms and results". In: *2016 IEEE 55th Conference on Decision and Control (CDC)*. IEEE. 2016, pp. 6795–6801.

[4] Artin Spiridonoff, Alex Olshevsky, and Ioannis Ch. Paschalidis. "Robust Asynchronous Stochastic Gradient-Push: Asymptotically Optimal and Network-Independent Performance for Strongly Convex Functions". In: *Journal of Machine Learning Research* 21.58 (2020), pp. 1–47.

[5] Ye Tian, Ying Sun, and Gesualdo Scutari. "Asy-sonata: Achieving linear convergence in distributed asynchronous multiagent optimization". In: *2018 56th Annual Allerton Conference on Communication, Control, and Computing (Allerton)*. IEEE. 2018, pp. 543–551.

[6] Tianyu Wu et al. "Decentralized consensus optimization with asynchrony and delays". In: *IEEE Transactions on Signal and Information Processing over Networks* 4.2 (2018), pp. 293–307.