SEQUENTIAL ESTIMATION OF LINEAR MODELS IN DISTRIBUTED SETTINGS

Yunlong Wang and Petar M.Djurić

Department of Electrical and Computer Engineering
Stony Brook University, Stony Brook, NY 11794
Email: {yunlong.wang,petar.djuric}@stonybrook.edu

ABSTRACT

In this paper, we consider the problem of distributed sequential estimation of time invariant parameters in a network of cooperative agents. We study a system where the agents quantify their respective beliefs in the unknown parameters by approximations of the posteriors of the parameters with multivariate Gaussians. At every time instant each agent carries out three operations, (a) it receives private measurements distorted by additive noise, (b) it exchanges information about its belief in the estimated parameters with its neighbors, and (c) it updates its belief with the new information. Since we consider distributed processing in the network, it is challenging to provide an optimal strategy where the agents update their believes using the Bayes' rule in every iteration. In this work, instead, we propose a method which does not process the data based on Bayes theory and yet allows the agents to reach asymptotically the optimal Bayesian belief held by a fictitious fusion center. We provide convergence analysis of the method and demonstrate its performance by simulations.

Index Terms— Distributed estimation, Bayesian estimation, consensus algorithms, cooperative agents,

1. INTRODUCTION

In this paper we propose a distributed algorithm for sequential estimation in a connected network of cooperative agents. In the network, any two neighboring agents have a perfect bidirectional communication. At each time instant, every agent receives private observations with information about time invariant parameters of a linear model. The observations are distorted by additive Gaussian noise. The agents quantify their beliefs in the unknown parameters by their respective posteriors of the parameters, which are approximated by multivariate Gaussian distributions. Since the agents are cooperative, they exchange information about their beliefs, and that allows them to modify their beliefs in the parameters. We propose an information fusion/diffusion strategy that permits the agents to reach asymptotically the same Bayesian belief held by a fictitious fusion center with access to all the measurements of the agents.

The problem of Bayesian learning in cooperative networks has been widely studied [1, 2, 3]. The objective there is that the agents asymptotically reach the performance of a Bayesian fusion center by cooperation. In this paper, we focus on distributed estimation. Various algorithms for distributed estimation have been proposed, some of them based on diffusion [4, 5] and others on consensus strategies [6, 7].

Our work can be viewed as an extension of [7] and [8] to the sequential scenario. The proposed method is an extension of the *running consensus* presented in [9], where the sequential averaging problem of scalars is considered. In comparison to the work in [9], this paper develops a method where every agent's belief is a function of two matrix statistics, both updated by the averaging consensus method. We analyze the performance of the method theoretically and demonstrate its performance by simulations.

The paper is organized as follows. In the next section we state the problem. In Section 3, we provide a brief review on consensus algorithms and present the proposed method. We prove that the method is unbiased and analyze its asymptotical property in Section 4. Simulation results are given in Section 5, and concluding remarks in Section 6.

2. PROBLEM STATEMENT

We consider distributed estimation in a network of cooperative agents A_i, $i \in \mathcal{N}_A = \{1, 2, ..., N\}$. The connections among the agents are described by an undirected graph $G = (\mathcal{N}_A, \mathcal{E})$, where $\mathcal{E}$ is the set of edges of the graph and where the agents A_i and A_j can directly exchange information if and only if $(i, j) \in \mathcal{E}$. In this paper, we assume that the topology of the network is time invariant and that the communication between agents is perfect.

We assume that $\forall i \in \mathcal{N}_A$, at any time instant $t \in \mathbb{N}^+$, A_i observes a vector of data $\mathbf{y}_i[t] \in \mathbb{R}^{M\times 1}$ generated by the following linear model:

$$\mathbf{y}_i[t] = \mathbf{H}_i[t]\boldsymbol{\theta} + \mathbf{w}_i[t], \quad (1)$$

This work was supported by NSF under Award CCF-1018323.

where $\boldsymbol{\theta} \in \mathbb{R}^{K\times 1}$ is a vector of unknown parameters of interest, and $\mathbf{w}_i[t]$ is observation noise modeled as a Gaussian random vector with zero-mean and covariance $\boldsymbol{\Sigma}_i$. We assume that $\mathbf{w}_i[t]$ is white in time and independent among different agents. It is also assumed that both $\mathbf{H}_i[t]$ and $\boldsymbol{\Sigma}_i$ represent private information known only to agent A_i. The agents model their prior belief about $\boldsymbol{\theta}$ by a noninformative prior and the posterior by a multivariate Gaussian. At every time instant, every agent exchanges quantities with its neighbors, which the agents use together with the private signals to update their beliefs in $\boldsymbol{\theta}$.

If there exists an ideal, centralized fusion center with access to all $\mathbf{y}_i[t]$, $\mathbf{H}_i[t]$ and $\boldsymbol{\Sigma}_i$, at time instant t its belief about $\boldsymbol{\theta}$ is the posterior of $\boldsymbol{\theta}$. By Gauss-Markov Theorem, this belief is a multivariate Gaussian distribution [10], i.e.,

$$\begin{aligned}\beta_{fc}[t] &= p(\boldsymbol{\theta}|\mathcal{I}_{fc}[t]) \\ &= \mathfrak{N}\left(\widehat{\boldsymbol{\theta}}_{fc}[t], \mathbf{C}_{fc}[t]\right),\end{aligned} \tag{2}$$

where $\mathfrak{N}(\cdot,\cdot)$ denotes a multivariate Gaussian distribution, the subscript fc emphasizes that the statistics belong to the fusion center, $\mathcal{I}_{fc}[t]$ represents all the information up to time instant t, which includes $\mathbf{y}_i[\tau]$, $\mathbf{H}_i[\tau]$ and $\boldsymbol{\Sigma}_i$, for each $i \in \mathcal{N}_A$ and $\tau \in \{1, 2, \cdots, t\}$, and where

$$\begin{aligned}\widehat{\boldsymbol{\theta}}_{fc}[t] &= \left(\sum_{\tau=1}^{t}\sum_{i=1}^{N}\mathbf{P}_i[\tau]\right)^{-1}\left(\sum_{\tau=1}^{t}\sum_{i=1}^{N}\mathbf{s}_i[\tau])\right) \\ &= \mathbf{D}_{fc}^{-1}[t]\boldsymbol{\eta}_{fc}[t],\end{aligned} \tag{3}$$

with $\mathbf{P}_i[\tau] \in \mathbb{R}^{K\times K}$ and $\mathbf{s}_i[\tau] \in \mathbb{R}^{K\times 1}$ being defined by

$$\begin{aligned}\mathbf{P}_i[\tau] &= \mathbf{H}_i[\tau]^\top\boldsymbol{\Sigma}_i^{-1}\mathbf{H}_i[\tau], \\ \mathbf{s}_i[\tau] &= \mathbf{H}_i[\tau]^\top\boldsymbol{\Sigma}_i^{-1}\mathbf{y}_i[\tau],\end{aligned}$$

and

$$\mathbf{D}_{fc}^{-1}[t] = \left(\sum_{\tau=1}^{t}\sum_{i=1}^{N}\mathbf{P}_i[\tau]\right)^{-1}, \tag{4}$$

$$\boldsymbol{\eta}_{fc}[t] = \sum_{\tau=1}^{t}\sum_{i=1}^{N}\mathbf{s}_i[\tau]. \tag{5}$$

The $K \times K$ covariance matrix $\mathbf{C}_{fc}[t]$ is given by,

$$\mathbf{C}_{fc}[t] = \mathbf{D}_{fc}^{-1}[t], \tag{6}$$

where the matrix $\mathbf{D}_{fc}[t]$ is referred to as the precision matrix.

3. DISTRIBUTED ESTIMATION

An examination of (3) reveals that in a distributed setting every agent will have the same belief as a fusion center if it can obtain the summation of all $\mathbf{P}_i[t]$s and $\mathbf{s}_i[t]$s. Hence, we cast the problem as one of distributed summation of sequential data. Note that the computation of (3) requires knowledge of the total number of agents in the network. This number, however, can be obtained by running a separate consensus algorithm. In this section we propose the method where agents can asymptotically achieve the belief held by the fusion fusion center by cooperation with neighboring agents.

We propose that at every t, each agent A_i keeps one matrix $\mathbf{D}_i[t] \in \mathbb{R}^{K\times K}$ and one vector $\boldsymbol{\eta}_i[t] \in \mathbb{R}^{K\times 1}$ to approximate the two statistics in (3) held by the fusion center, i.e., $\mathbf{D}_i[t] \approx \mathbf{D}_{fc}[t]$ and $\boldsymbol{\eta}_i[t] \approx \boldsymbol{\eta}_{fc}[t]$.

At $t = 0$, all the elements of $\mathbf{D}_i[0]$ and $\boldsymbol{\eta}_i[0]$ are initialized to be zero. Let $q_{ij} \in \mathbb{R}$ be a non-negative weight that agent A_i assigns to its neighbor A_j. Then with the matrix $\mathbf{Q} \in \mathbb{R}^{N\times N}$ we represent all the weights of all the agents in the network. At time t, any A_i, $i \in \mathcal{N}_A$, after exchanging information with its neighbors, updates its statistics by

$$\begin{aligned}\mathbf{D}_i[t] &= \sum_{j=1}^{N} q_{i,j}\mathbf{D}_j[t-1] + N\mathbf{P}_i[t] \\ &= N\sum_{\tau=1}^{t}\sum_{j=1}^{N}\phi_{i,j}[t-\tau]\mathbf{P}_j[\tau],\end{aligned} \tag{7}$$

$$\begin{aligned}\boldsymbol{\eta}_i[t] &= \sum_{j=1}^{N} q_{i,j}\boldsymbol{\eta}_j[t-1] + N\mathbf{s}_i[t] \\ &= N\sum_{\tau=1}^{t}\sum_{j=1}^{N}\phi_{i,j}[t-\tau]\mathbf{s}_j[\tau],\end{aligned} \tag{8}$$

where $\phi_{i,j}[t] \in \mathbb{R}$ denotes the element in the ith row and jth column of $\mathbf{Q}^t$. We propose that A_i forms its belief $\beta_i[t]$ about $\boldsymbol{\theta}$ as a multivariate Gaussian distribution, i.e., $\beta_i[t] \sim \mathfrak{N}(\widehat{\boldsymbol{\theta}}_i[t], \mathbf{C}_i[t])$, where

$$\widehat{\boldsymbol{\theta}}_i[t] = \mathbf{D}_i^{-1}[t]\boldsymbol{\eta}_i[t], \quad \forall i \in \mathcal{N}_A, \ \forall t \in \mathbb{N}^+, \tag{9}$$

and

$$\mathbf{C}_i[t] = \mathbf{D}_i^{-1}[t], \quad \forall i \in \mathcal{N}_A, \ \forall t \in \mathbb{N}^+. \tag{10}$$

We remark that if $\phi_{i,j}[t] = \dfrac{1}{N}$, $\forall i, j \in \mathcal{N}_A$ and $\forall t \in \mathbb{N}^+$, every agent will perform as well as the fusion center. Aiming at achieving this goal, we require that the weight matrix $\mathbf{Q}$ satisfies the following three conditions:

$$\mathbf{1}^\top\mathbf{Q} = \mathbf{1}^\top, \quad \mathbf{Q}\mathbf{1} = \mathbf{1}, \quad \rho\left(\mathbf{Q} - (1/N)\mathbf{1}\mathbf{1}^\top\right) < 1, \tag{11}$$

where $\mathbf{1}$ denotes an $N \times 1$ column vector with all of its entries equal to one, and $\rho(\cdot)$ is the spectral radius of the matrix inside the parentheses.

According to [11], it holds that

$$\lim_{t\to\infty} \mathbf{Q}^t = \frac{1}{N}\mathbf{1}\mathbf{1}^\top, \tag{12}$$

and

$$\lim_{t\to\infty} \phi_{i,j}[t] = \frac{1}{N}, \quad \forall i,j \in \mathcal{N}_A. \tag{13}$$

Throughout this paper, we assume that the matrix $\mathbf{Q}$ is symmetric and that it satisfies the three conditions in (11).

4. ANALYSIS

In this section, we first examine the bias of the estimates of the proposed method. The following theorem maintains that the estimates of the agents are unbiased.

Theorem 1 *Under the conditions listed in the statement of the problem, $\forall i \in \mathcal{N}_A$ and $\forall t \in \mathbb{N}^+$, the estimate in* (9) *held by agent A_i at time t is unbiased, i.e.,*

$$\mathbb{E}\,\widehat{\boldsymbol{\theta}}_i[t] = \boldsymbol{\theta}. \tag{14}$$

Proof: The theorem can be proved by mathematical induction. At time $t=1$, $\forall i \in \mathcal{N}_A$, we have

$$\begin{aligned}\mathbb{E}\,\widehat{\boldsymbol{\theta}}_i[1] &= \mathbb{E}\left(\mathbf{D}_i^{-1}[1]\boldsymbol{\eta}_i[1]\right) \\ &= \mathbb{E}\left((N\mathbf{C}_i^{-1}[1])^{-1}(N\mathbf{C}_i^{-1}[1]\widetilde{\boldsymbol{\theta}}_i[1])\right) \\ &= \mathbb{E}\,\widetilde{\boldsymbol{\theta}}_i[1] = \boldsymbol{\theta}.\end{aligned} \tag{15}$$

Assuming that at time t we have $\mathbb{E}\,\widehat{\boldsymbol{\theta}}_i[t] = \boldsymbol{\theta}$, $\forall i \in \mathcal{N}_A$, then at time $t+1$,

$$\begin{aligned}&\mathbb{E}\,\widehat{\boldsymbol{\theta}}_i[t+1] = \mathbb{E}\left(\mathbf{D}_i^{-1}[t+1]\boldsymbol{\eta}_i[t+1]\right) \\ &= \mathbf{D}_i^{-1}[t+1]\left(\textstyle\sum_{j=1}^N q_{i,j}\mathbf{D}_j[t]\mathbb{E}\left(\widehat{\boldsymbol{\theta}}_j[t]\right)\right. \\ &\quad \left.+ N\mathbf{P}_i[t+1]\mathbb{E}\left(\mathbf{P}_i^{-1}[t+1]\mathbf{s}_i[t+1]\right)\right) \\ &= \mathbf{D}_i^{-1}[t+1]\Big(\textstyle\sum_{j=1}^N q_{i,j}\mathbf{D}_j[t] + N\mathbf{P}_i[t+1]\Big)\boldsymbol{\theta} \\ &= \mathbf{D}_i^{-1}[t+1]\mathbf{D}_i[t+1]\boldsymbol{\theta} = \boldsymbol{\theta},\end{aligned} \tag{16}$$

where the second equality is due to the fact that $\boldsymbol{\eta}_j[t] = \mathbf{D}_j[t]\widehat{\boldsymbol{\theta}}_j[t]$, the third equality is because the local estimates are unbiased (i.e., $\mathbb{E}\,\mathbf{P}_i^{-1}[t+1]\mathbf{s}_i[t+1] = \boldsymbol{\theta}$), and we assume $\mathbb{E}\,\widehat{\boldsymbol{\theta}}_j[t] = \boldsymbol{\theta}$ in the inductive step. □

With the second theorem, we show that the belief of every agent will converge to the global optimal obtained by a fictitious fusion center.

Theorem 2 *Let the conditions listed in the statement of the problem hold and let $\boldsymbol{\Sigma}_i$ and $\mathbf{H}_i[t]$ be bounded $\forall i \in \mathcal{N}_A$. Then by the proposed method, the Kullback-Leibler (KL) divergence between any agent's belief $\beta_i[t]$ and the belief of the fictitious fusion center $\beta_{fc}[t]$ converges to zero as time goes to infinity, i.e.,*

$$\lim_{t\to\infty} D_{KL}\left(\beta_i[t]\middle\|\beta_{fc}[t]\right) = 0, \quad \forall i \in \mathcal{N}_A. \tag{17}$$

Proof: According to the definition of KL-divergence between two multivariate Gaussian distributions, we have that the sufficient conditions for (17) are given by,

$$\lim_{t\to\infty}\left(\widehat{\boldsymbol{\theta}}_{fc}[t] - \widehat{\boldsymbol{\theta}}_i[t]\right) = \mathbf{0} \quad \forall i \in \mathcal{N}_A, \tag{18}$$

and

$$\lim_{t\to\infty} \mathbf{D}_{fc}[t]\mathbf{D}_i^{-1}[t] = \mathbf{I} \quad \forall i \in \mathcal{N}_A, \tag{19}$$

where $\mathbf{0} \in \mathbb{R}^{K\times 1}$ is a vector of K zeros, and $\mathbf{I}$ is the $K \times K$ identity matrix.

By (9), we rewrite the estimates held by agent A_i at time instant t as,

$$\begin{aligned}\widehat{\boldsymbol{\theta}}_i[t] &= \mathbf{D}_i^{-1}[t]\boldsymbol{\eta}_i[t] \\ &= \left(\mathbf{D}_i[t] - \mathbf{D}_{fc}[t] + \mathbf{D}_{fc}[t]\right)^{-1} \\ &\quad \times \left(\boldsymbol{\eta}_i[t] - \boldsymbol{\eta}_{fc}[t] + \boldsymbol{\eta}_{fc}[t]\right) \\ &= (\mathbf{E}_i[t] + \mathbf{D}_{fc}[t])^{-1}(\boldsymbol{\epsilon}_i[t] + \boldsymbol{\eta}_{fc}[t]),\end{aligned} \tag{20}$$

where $\mathbf{E}_i[t] = \mathbf{D}_i[t] - \mathbf{D}_{fc}[t]$ and $\boldsymbol{\epsilon}_i[t] = \boldsymbol{\eta}_i[t] - \boldsymbol{\eta}_{fc}[t]$. By using similar reasoning, we write,

$$\mathbf{D}_{fc}[t]\mathbf{D}_i^{-1}[t] = \mathbf{D}_{fc}[t](\mathbf{E}_i[t] + \mathbf{D}_{fc}[t])^{-1}. \tag{21}$$

With the assumption that $\boldsymbol{\Sigma}_i[t]$ and $\mathbf{H}_i[t]$ are bounded $\forall i \in \mathcal{N}_A$, we have that all the covariance matrices $\mathbf{C}_i[t]$ are bounded and positive definite. Therefore, we have that $\mathbf{D}_{fc}[t] = \sum_{\tau=1}^t \sum_{i=1}^N \mathbf{C}_i^{-1}[\tau]$ is unbounded as t goes to infinity. Also, note that the estimate $\widehat{\boldsymbol{\theta}}_{fc}[t]$ in (3) converges to $\boldsymbol{\theta}$ as t goes to infinity, and therefore it holds that $\boldsymbol{\eta}_{fc}[t]$ is also unbounded when time t goes to infinity. As indicated in (20), since $\widehat{\boldsymbol{\theta}}_{fc}[t] = \mathbf{D}_{fc}^{-1}[t]\boldsymbol{\eta}_{fc}[t]$, it is sufficient to show that $\forall i \in \mathcal{N}_A$, both $\mathbf{E}_i[t]$ and $\boldsymbol{\epsilon}_i[t]$ are bounded to get (18) and (19).

By (3), (7), and (8), we have,

$$\begin{aligned}\|\mathbf{E}_i[t]\| &= \left\|\sum_{\tau=1}^t\sum_{j=1}^N \left(N\phi_{i,j}[t-\tau]\mathbf{P}_j[\tau] - \mathbf{P}_j[\tau]\right)\right\| \\ &\leq \left(\sum_{\tau=1}^t\sum_{j=1}^N \phi_{i,j}[t-\tau] - t\right) N\mathcal{P}[t],\end{aligned} \tag{22}$$

and

$$\begin{aligned}\|\boldsymbol{\epsilon}_i[t]\| &= \left\|\sum_{\tau=1}^t\sum_{j=1}^N \left(N\phi_{i,j}[t-\tau]\mathbf{s}_j[\tau] - \mathbf{s}_j[\tau]\right)\right\| \\ &\leq \left(\sum_{\tau=1}^t\sum_{j=1}^N \phi_{i,j}[t-\tau] - t\right) N\mathcal{T}[t],\end{aligned} \tag{23}$$

where $\|\cdot\|$ denotes the Euclidean-norm, and

$$\begin{aligned}\mathcal{P}[t] &= \max_{j\in\mathcal{N}_A,\tau\in\{1,\cdots,t\}} \|\mathbf{P}_j[\tau]\|\,,\\ \mathcal{T}[t] &= \max_{j\in\mathcal{N}_A,\tau\in\{1,\cdots,t\}} \|\mathbf{s}_j[\tau]\|\,.\end{aligned}$$

Since we have that both $\mathcal{C}[t]$ and $\mathcal{T}[t]$ are bounded, we next show that $g_i[t] = \sum_{\tau=1}^{t}\sum_{j=1}^{N}\phi_{i,j}[t-\tau] - t$ is bounded for all $i \in \mathcal{N}_A$. We construct a vector $\mathbf{g}[t]$ with elements $g_i[t]$ as follows:

$$\mathbf{g}[t] = \sum_{\tau=1}^{t}\mathbf{Q}^{t-\tau}\mathbf{1} - \sum_{\tau=1}^{t}\frac{\mathbf{1}\mathbf{1}^\top}{N}\mathbf{1}, \tag{24}$$

where $\mathbf{1} \in \mathbb{R}^{N\times 1}$ denotes a vector of ones. By the properties of the 2-norm, we have

$$\begin{aligned}\|\mathbf{g}[t]\| &= \left\|\sum_{\tau=1}^{t}\mathbf{Q}^{t-\tau}\mathbf{1} - \sum_{\tau=1}^{t}\frac{\mathbf{1}\mathbf{1}^\top}{N}\mathbf{1}\right\|\\ &= \left\|\sum_{\tau=1}^{t}(\mathbf{Q}-\frac{\mathbf{1}\mathbf{1}^\top}{N})^{t-\tau}\mathbf{1}\right\|\\ &\le \sum_{\tau=1}^{t}\left\|(\mathbf{Q}-\frac{\mathbf{1}\mathbf{1}^\top}{N})\right\|^{\,t-\tau}\left\|\mathbf{1}\right\|,\end{aligned}\tag{25}$$

where the second equality is due to the fact that $\mathbf{Q}$ is doubly stochastic and the matrix $\mathbf{I} - \frac{\mathbf{1}\mathbf{1}^\top}{N}$ is a projection matrix and the relation $\le$ is due to the property of the Euclidean-norm. Since $\mathbf{Q}$ is assumed to be symmetric, then $\mathbf{Q} - \frac{\mathbf{1}\mathbf{1}^\top}{N}$ is also a symmetric matrix whose spectral radius is equal to its 2-norm. Hence we have,

$$\|\mathbf{g}[t]\| \le \sqrt{N}\sum_{\tau=1}^{t}\rho\big(\mathbf{Q}-\frac{\mathbf{1}\mathbf{1}^\top}{N}\big)^{t-\tau}. \tag{26}$$

Noting that $\rho\big(\mathbf{Q}-\frac{\mathbf{1}\mathbf{1}^\top}{N}\big) < 1$, we have proved that the Euclidean-norm of $\mathbf{s}[t]$ is bounded as t goes to infinity, which shows that every element $g_i[t]$ is also bounded $\forall i \in \mathcal{N}_A$ and $\forall t \in \mathbb{N}^+$. With this, the proof of (18) and (19) is completed, and thus, the main claim of the theorem with (17). □

5. SIMULATION RESULTS

In this section, we present simulations of the proposed algorithm and numerical comparisons between the convergence rate of agents with different number of neighbors. In all the experiments, the multi-agent system was modeled as a random geometric graph $G(\mathcal{N}_A, \mathcal{E})$ [12], where the N agents were chosen uniformly and independently on a square of size 1×1. Each pair was connected if the Euclidean distance between the nodes was smaller than $r(N)$, where $r(N) = \sqrt{\frac{\log(N)}{N}}$ due to connectivity requirement.

In all the experiments, we set $\mathbf{Q}$ to be the updating matrix of the average consensus algorithm [13], which has the following form:

$$\mathbf{Q} = \mathbf{I} - \epsilon\mathbf{L}, \tag{27}$$

where $\mathbf{I} \in \mathbb{R}^{N\times N}$ is the identity matrix, $\mathbf{L}$ is the Laplacian matrix of the random graph G, and $\epsilon \in \mathbb{R}$ is a coefficient satisfying $\epsilon < 1/\max_i(deg(i))$, $\forall i \in \mathcal{N}_A$, with $deg(i)$ denoting the degree of node i. Since $\rho\big(\mathbf{Q} - \frac{\mathbf{1}\mathbf{1}^\top}{N}\big)$ has a smaller value when ϵ is larger, it can be expected that a larger ϵ results in a faster convergence.

In the first experiment, $N = 20$, $M = 5$, $K = 4$, $t \in \{1, 2, \cdots, 100\}$ and $\forall i \in \mathcal{N}_A$, $\forall t \in \mathbb{N}^+$, we set the elements of $\mathbf{H}_i[t] \in \mathbb{R}^{5\times 4}$ to be independent random variables uniformly distributed on $[2, 4]$, and set $\boldsymbol{\theta} \in \mathbb{R}^{K\times 1}$ to be normalized and whose elements were independent and uniformly distributed. Also, $\forall i \in \mathcal{N}_A$ we set $\boldsymbol{\Sigma}_i$ to be a random $M \times M$ matrix generated by a Wishart distribution, where the scale matrix was the identity matrix $\mathbf{I} \in \mathbb{R}^{5\times 5}$ and the degrees of freedom were 15. As a performance metric for the different methods, we used the mean KL-divergence at time instant t ($\overline{D}_{KL}[t]$) defined as the average value of $\sum_{i=1}^{N} D_{KL}\Big(\beta_i[t]\Big|\Big|\beta_{fc}[t]\Big)/N$ over 500 implementations.

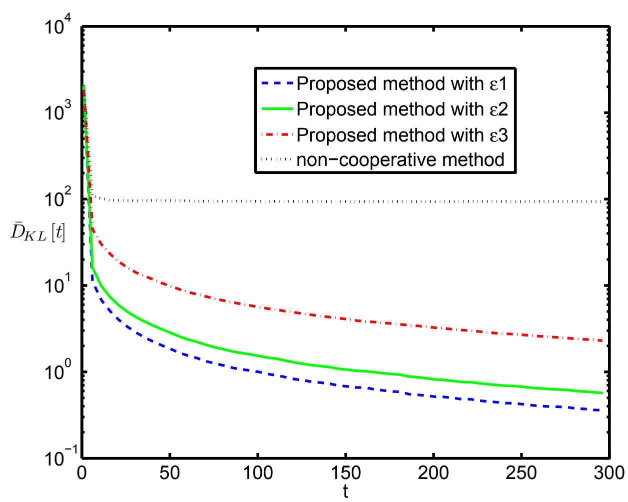

Fig. 1. Asymptotical performance of the proposed method.

In this experiment, the network of agents implemented the proposed method for 500 times with four different ϵs. More specifically, we set $\epsilon_1 = 0.9/\max_i(deg(i))$, $\epsilon_2 = 0.5/\max_i(deg(i))$, $\epsilon_3 = 0.1/\max_i(deg(i))$, and $\epsilon_4 = 0$. We recall that the case of $\epsilon = 0$ corresponds to a non-cooperative scheme. In each implementation, we recorded the mean KL-divergence between the beliefs of agents and that of the fictitious fusion center at every t. In Fig. 1, we plotted the $\overline{D}_{KL}[t]$ of the proposed method with different ϵs, where the y-axis shows the $\overline{D}_{KL}[t]$ in log scale and the x-axis represents time from 1 to 300. It can be seen that the mean KL-divergence of our method for all positive ϵs converges to zero with time, whereas the one related to the noncooperative method, does not. Also, we can see that larger values of ϵ yield faster convergence.

In the second experiment, we compared the performance of agents with different degrees when they implement the proposed method. In the random network we chose an agent, referred to as A_1, with degree 5, another agent, A_2, with degree 9, and a third agent, A_3, with degree 1. At each time instant t, every agent receives the information generated by the linear model described in the experiment one and it implements the proposed method using the update matrix Q with ϵ_1. For comparisons, we defined the performance metric $\overline{D}^i_{KL}[t]$ by the average value of $D_{KL}\left(\beta_i[t]\middle|\middle|\beta_{fc}[t]\right)/N$ over 500 trials. In Fig. 2., we plotted the evolution of $\overline{D}^i_{KL}[t]$ of these three agents respectively, where it shows that the agent with the largest degree has the fastest convergence.

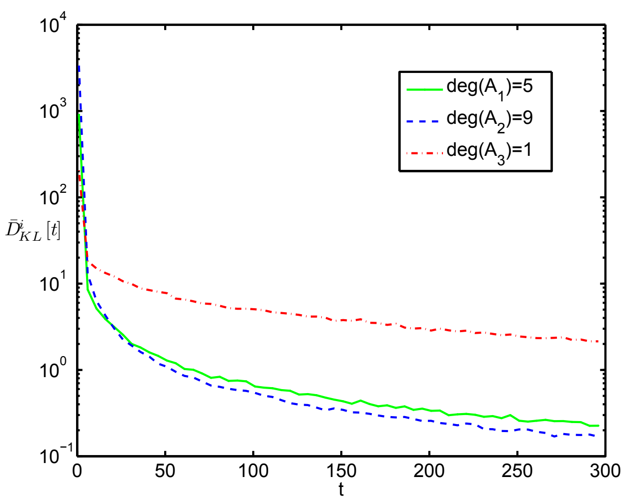

Fig. 2. Comparison between the evolution of belief held by agents with different degrees.

6. CONCLUSION

An important challenge for cooperative agents is efficient processing of sequential private information and information received from their neighbors. In this paper we presented one approach of addressing this problem when the observations of the agents are generated by a Gaussian linear model and the objective is that the agents reach the posterior about the linear parameters of the model given the global information. We proved that with the proposed method, the estimates held by the agents are unbiased. We also showed that the belief of every agent asymptotically reaches the Bayesian belief of a fusion center. The computer simulations demonstrated that the agents had performance that is predicted by the theoretical analysis of the method.

7. REFERENCES

[1] P. M. Djurić and Y. Wang, "Distributed Bayesian learning in multiagent systems: Improving our understanding of its capabilities and limitations," *IEEE Signal Processing Magazine*, vol. 29, no. 2, pp. 65–76, 2012.

[2] D. Gale and S. Kariv, "Bayesian learning in social networks," *Games and Economic Behavior*, vol. 45, no. 2, pp. 329–346, 2003.

[3] E. Mossel and O. Tamuz, "Efficient Bayesian learning in social networks with Gaussian estimators," *arXiv preprint arXiv:1002.0747*, 2010.

[4] F. S. Cattivelli, C. G. Lopes, and A. H. Sayed, "Diffusion recursive least-squares for distributed estimation over adaptive networks," *IEEE Transactions on Signal Processing*, vol. 56, no. 5, pp. 1865–1877, 2008.

[5] C. G. Lopes and A. H. Sayed, "Diffusion least-mean squares over adaptive networks: Formulation and performance analysis," *IEEE Transactions on Signal Processing*, vol. 56, no. 7, pp. 3122–3136, 2008.

[6] I. D. Schizas, A. Ribeiro, and G. B. Giannakis, "Consensus in ad hoc WSNs with noisy linkspart I: Distributed estimation of deterministic signals," *IEEE Transactions on Signal Processing*, vol. 56, no. 1, pp. 350–364, 2008.

[7] L. Xiao, S. Boyd, and S. Lall, "A scheme for robust distributed sensor fusion based on average consensus," in *Information Processing in Sensor Networks, 2005. IPSN 2005. Fourth International Symposium on*. IEEE, 2005, pp. 63–70.

[8] R. Olfati-Saber, J. A. Fax, and R. M. Murray, "Consensus and cooperation in networked multi-agent systems," *Proceedings of the IEEE*, vol. 95, no. 1, pp. 215–233, 2007.

[9] P. Braca, S. Marano, and V. Matta, "Enforcing consensus while monitoring the environment in wireless sensor networks," *IEEE Transactions on Signal Processing*, vol. 56, no. 7, pp. 3375–3380, 2008.

[10] S. M. Kay, *Fundamentals of Statistical Signal Processing, Volume I: Estimation Theory (v. 1)*, Prentice Hall, 1993.

[11] L. Xiao and S. Boyd, "Fast linear iterations for distributed averaging," *Systems & Control Letters*, vol. 53, no. 1, pp. 65–78, 2004.

[12] M. Penrose, *Random Geometric Graphs*, vol. 5, Oxford University Press Oxford, UK:, 2003.

[13] R. Olfati-Saber and R. M. Murray, "Consensus problems in networks of agents with switching topology and time-delays," *IEEE Transactions on Automatic Control*, vol. 49, no. 9, pp. 1520–1533, 2004.