Distributed Sequential Estimation in a Network of Cooperative Agents

Petar M. Djurić

with Yunlong Wang

Department of Electrical and Computer Engineering
Stony Brook University
Stony Brook, NY 11794, USA

February 18, 2013

Outline

Outline

Interest

- We are interested in Bayesian learning in a network of cooperative agents.
- The agents exchange information with their neighbors only.
- We aim at finding methods that asymptotically have the performance as a Bayesian fusion center.
- In general, we want to find the minimal information that the agents need to exchange so that their performance gets as close as possible to the performance of the fusion center (not discussed here).

State-of-art

- This problem has been addressed by using consensus and diffusion strategies.
- Average consensus and gossip algorithms have been studied extensively in recent years, especially in the control literature.
- These strategies have been applied to various types of problems including multi-agent formations, distributed optimization, distributed control, distributed detection, and distributed estimation.

State-of-art (cont.)

- Original implementations of the consensus strategies required the use of two time scales: one for the acquisition of measurements and the other for the consensus.
- More recent work on consensus-based methods is on single time scales.
- As alternatives to the consensus method, diffusion methods have been proposed which inherently have single time scale implementations.
- It has been shown that the dynamics of the consensus and diffusion strategies differ in important ways.

State-of-art (cont.)

- More specifically, recently, Sayed and his group published a paper (IEEE Transactions on Signal Processing, Dec. 2012) in which they proposed two types of diffusion strategies for distributed estimation.
- They are termed ATC (adapt-then-combine) and CTA (combine-then-adapt) strategies.
- They studied the properties of these strategies on a linear regression problem and compared them to the consensus-based strategy.

State-of-art (cont.)

- ATC method: At every time instant t, after receiving private signals, the agents update their estimates using the received signals, and then combine them with the estimates from their neighbors. They broadcast the obtained estimates.
- CTA method: At every time instant t, the agents first combine all the estimates, and then update the so obtained estimates using the received signals. They broadcast the obtained estimates.

State-of-art (cont.)

- They have found that the ATC has the best and that the consensus method, the worst properties.
- More specifically, the diffusion strategies have lower mean square deviation than consensus methods, and their mean-square deviation is insensitive to the choice of the combination weights.

Outline

The setup

- A network of N cooperative agents aim at estimating a vector of time invariant parameters.
- They are spatially distributed and linked together through a connected topology.
- The communication among two neighboring agents is bidirectional.
- The agents receive private signals which are modeled by a linear model with the same fixed parameters.

The network model

We consider a distributed estimation in a network of cooperative agents A_i, $i \in \mathcal{N}_A = \{1, 2, ..., N\}$:

- $G = (\mathcal{N}_A, \mathcal{E})$ is a graph that describes the connections among the agents.
- A_i and A_j can directly exchange information if and only if $\{i, j\} \in \mathcal{E}$.
- We assume that the topology of the network is time invariant and that the communication between any two communicating agents is perfect.

The observation model

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

$$\mathbf{y}_i[t] = \mathbf{H}_i[t]\boldsymbol{\theta} + \mathbf{w}_i[t].$$

- In this work, $\boldsymbol{\theta} \in \mathbb{R}^{K\times 1}$ is a vector of unknown parameters to be estimated.
- The observation noise $\mathbf{w}_i[t]$ is an independent random vector from previous and future time instants with zero-mean and covariance $\boldsymbol{\Sigma}_i[t]$.
- We assume that the $\mathbf{w}_i[t]$s are independent among different agents.
- Both $\mathbf{H}_i[t]$ and $\boldsymbol{\Sigma}_i[t]$ represent private information known *only* to the agent A_i.

The local LMMSE

At any time instant $t \in \mathbb{N}^+$, the LMMSE estimate of A_i from its own data, for any $i \in \mathcal{N}_A$, is

$$\widetilde{\boldsymbol{\theta}}_i[t] = \left(\mathbf{H}_i[t]^\top \boldsymbol{\Sigma}_i[t]^{-1}\mathbf{H}_i[t]\right)^{-1}\mathbf{H}_i[t]^\top \boldsymbol{\Sigma}_i[t]^{-1}\mathbf{y}_i[t].$$

- We refer to this estimate as the *local estimate* from the private signals.
- We assume that $\forall i \in \mathcal{N}_A$ and $\forall t \in \mathbb{N}^+$, the matrix $\mathbf{H}_i[t]^\top \boldsymbol{\Sigma}_i[t]^{-1}\mathbf{H}_i[t]$ has full rank.
- The covariance of the estimate is given by,

$$\mathbf{C}_i[t] = (\mathbf{H}_i[t]^\top \boldsymbol{\Sigma}_i[t]^{-1}\mathbf{H}_i[t])^{-1}.$$

The LMMSE estimate of a fictitious fusion center

The LMMSE of a fictitious fusion center is given by,

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

- $\mathbf{C}_{fc}[t] = \left(\sum_{\tau=1}^{t}\sum_{i=1}^{N}\mathbf{C}_i^{-1}[\tau]\right)^{-1}$ is the covariance of $\widehat{\boldsymbol{\theta}}_{fc}[t]$.
- $\boldsymbol{\eta}_{fc}[t] = \sum_{j=1}^{t}\sum_{i=1}^{N}\left(\mathbf{C}_i^{-1}[t]\widetilde{\boldsymbol{\theta}}_{i,t}\right)$.
- The suffix *fc* here emphasizes that the statistics are of the fusion center.

Objective

- All the agents' estimates asymptotically reach the estimate of the fictitious fusion center, and
- In meeting the objective, we want to have as little communication between the agents as possible.

Outline

The concept

- Every agent will have the same performance as the fusion center if it can obtain the summation of all $\mathbf{C}_i^{-1}[t]$s and $\mathbf{C}_i^{-1}[t]\widetilde{\theta}_i[t]$s.
- Hence, we can cast this problem as a problem of distributed summation of sequential data.
- The agents' performance would be optimal if they know the average values of the above two statistics.

The algorithm in a scalar form

Consider a system where at t ($t \in \mathbb{N}^+$) each agent A_i collects a new measurement, which is denoted by $x_i[t] \in \mathbb{R}$. We also denote by $s_i[t]$ the state of agent A_i at time t. Then the state of the agent A_i is updated as follows:

$$s_i[t] = \sum_{j=1}^{N} q_{ij} s_j[t-1] + N x_i[t].$$

- $s_i[0] = 0$, $\forall i \in \mathcal{N}_A$.
- q_{ij} is a nonnegative weight, which agent A_i assigns to its neighbor A_j.
- $q_{ij} = 0$ if $\{i, j\} \notin \mathcal{E}$.

The algorithm in a vector form

In a vector form, the update rule is given by,

$$\begin{aligned} \mathbf{s}[t] &= \mathbf{Q}\mathbf{s}[t-1] + N\mathbf{x}[t] \\ &= N\left(\mathbf{Q}^{t-1}\mathbf{x}[1] + \mathbf{Q}^{t-2}\mathbf{x}[2] + \cdots + \mathbf{x}[t]\right). \end{aligned}$$

- $\mathbf{Q}$ is a matrix whose elements are q_{ij}.
- $\mathbf{s}[t]$ and $\mathbf{x}[t]$ are column vectors whose entries are the states and the measurements of the agents at t, respectively.
- We require that the matrix $\mathbf{Q}$ of non-negative weights satisfies the following three conditions ($\mathbf{1}$ denotes a $N \times 1$ column vector with entries equal to one):

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

  where $\rho(\cdot)$ denotes the spectral radius of the argument.

The algorithm in a vector form (cont.)

Then it holds that

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

- It is shown that the state of an agent is a summation of t approximations of $\sum_{i=1}^{N} x_i[\tau]$, $\tau = \{1, 2, \cdots, t\}$.
- Due to the asymptotical property of $\mathbf{Q}$, these t approximations will become more and more accurate as t grows.
- Earlier local statistics reach all the agents.

The method

- 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.
- They are $\mathbf{D}_i[t] \approx (\mathbf{C}_{fc}[t])^{-1}$ and $\boldsymbol{\eta}_i[t] \approx \boldsymbol{\eta}_{fc}[t]$.
- At $t = 0$, all elements of $\mathbf{D}_i[0]$ and $\boldsymbol{\eta}_i[0]$ are initialized to be zero.

The method (cont.)

Then at time t, for any agent $i \in \mathcal{N}_A$, agent A_i and its neighbors exchange information and update their statistics in the following form:

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

where $\phi_{i,j}[t]$ denotes the element at the ith row and jth column of $\mathbf{Q}^t$, and $\lim_{t\to\infty}\phi_{i,j}[t] = \frac{1}{N}$.

The method (cont.)

We also have

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

Then at time t, the estimate of $\boldsymbol{\theta}$ held by agent A_i, $i \in \mathcal{N}_A$, is given by

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

The algorithm

At time $t \in \mathbb{N}^+$, for any agent $i \in \mathcal{N}_A$, agent A_i carries out the following steps:

Step 1: Receives noisy observations and calculates the local estimates and their covariance,

$$\widetilde{\boldsymbol{\theta}}_i[t] = \left(\mathbf{H}_i[t]^\top \boldsymbol{\Sigma}_i[t]^{-1} \mathbf{H}_i[t]\right)^{-1} \mathbf{H}_i[t]^\top \boldsymbol{\Sigma}_i[t]^{-1} \mathbf{y}_i[t],$$

$$\mathbf{C}_i[t] = (\mathbf{H}_i[t]^\top \boldsymbol{\Sigma}_i[t]^{-1} \mathbf{H}_i[t])^{-1}.$$

Algorithm(cont.)

Step 2: Updates its states according to,

$$
\begin{aligned}
\mathbf{D}_i[t] &= \sum_{j=1}^{N} q_{i,j}\mathbf{D}_j[t-1] + N\mathbf{C}_i^{-1}[t] \\
\boldsymbol{\eta}_i[t] &= \sum_{j=1}^{N} q_{i,j}\boldsymbol{\eta}_j[t-1] + N\mathbf{C}_i^{-1}[t]\widetilde{\boldsymbol{\theta}}_i[t].
\end{aligned}
$$

Step 3: Exchanges its current states $\mathbf{D}_i[t]$ and $\boldsymbol{\eta}_i[t]$ with its neighbors.

Step 4: Updates its estimates of $\boldsymbol{\theta}$ from its states by

$$
\widehat{\boldsymbol{\theta}}_i[t] = \mathbf{D}_i[t]^{-1}\boldsymbol{\eta}_i[t].
$$

Outline

Unbiasedness

Theorem 1 *$\forall i \in \mathcal{N}_A$ and $\forall t \in \mathbb{N}^+$, the estimate $\widehat{\boldsymbol{\theta}}_i[t]$ held by agent A_i at time t is unbiased, i.e.*

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

The theorem is proved by mathematical induction.

Convergence of the first moment

Theorem 2 *Assuming that* $\mathbf{y}_i[t]$ *and* $\mathbf{H}_i[t]$ *are bounded, then by the proposed method,* $\forall i \in \mathcal{N}_A$ *the estimate held by agent* A_i *asymptotically converges to the estimate held by a fictitious fusion center, i.e.,*

$$\lim_{t\to\infty} \left(\widehat{\boldsymbol{\theta}}_i[t] - \widehat{\boldsymbol{\theta}}_{fc}[t]\right) = \mathbf{0}_{K\times 1} \quad \forall i \in \mathcal{N}_A,$$

where $\mathbf{0}_{K\times 1} \in \mathbb{R}^{K\times 1}$ *is a vector of zeros.*

There are two ways to prove the theorem.

- The first approach is based on the the convergence of the first moment. The estimate is a multiplication of two parts, and in both parts, the difference between the result from the fusion center and from the agent is bounded, while the similarity is unbounded. One can show that the ratio of the difference and the similarity is vanishing with t.
- The second approach is based on the comparison of the covariance matrices of the estimates of an agent and the fusion center.

Convergence of the second moment

Theorem 3 *In the case where the covariance matrices of the local estimates are identical, i.e.,* $\mathbf{C}_{i,t} = \mathbf{C}$, $\forall i \in \mathcal{N}_A$, $\forall t \in \mathbb{N}^+$, *every agent will asymptotically perform as well as the fusion center in the sense that its covariance matrix of the estimate,* $\mathbf{M}_i[t]$, *satisfies*

$$\lim_{t\to\infty}\left(\mathbf{M}_i[t] - \frac{\mathbf{C}}{Nt}\right) = \mathbf{0}_{K\times K} \quad \forall \in \mathcal{N}_A,$$

where $\mathbf{C}/(Nt)$ *is the covariance matrix of the estimate held by the fusion center, and* $\mathbf{0}_{K\times K} \in \mathbb{R}^{K\times K}$ *is a matrix of zeros.*

Outline

Topology

- The system was modeled as a random geometric graph $G(\mathcal{N}_A, \mathcal{E})$.
- The N agents were chosen uniformly and independently on a square of size 1×1.
- Each pair was connected if the Euclidian distance between the nodes was smaller than $r(N)$, where $r(N) = \sqrt{\frac{\log(N)}{N}}$ due to connectivity requirement.

The updating matrix

In all the experiments, we set $\mathbf{Q}$ to have the following form:

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

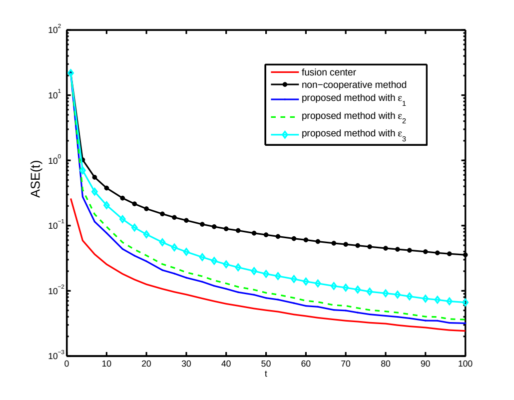
- $\mathbf{I} \in \mathbb{R}^{N \times N}$ is the identity matrix.
- $\mathbf{L}$ is the Laplacian matrix of the random graph G.
- $\epsilon \in \mathbb{R}$ is a coefficient satisfying $\epsilon < 1/\max_i(deg(i)), \forall i \in \mathcal{N}_A$, with $deg(i)$ denoting the degree of node i.
- Larger ε results in a faster convergence.

Parameter setting

- In the experiments, $\boldsymbol{\theta} = [4, 3, 2, 1]^\top$, $N = 15$, $M = 5$, $K = 4$, $t \in \{1, 2, \cdots, 100\}$.
- The elements of $\mathbf{H}_i[t] \in \mathbb{R}^{5\times 4}$ were random variables uniformly distributed on $[0, 3]$ and were independent from each other.
- $\boldsymbol{\Sigma}_i[t]$ was a diagonal $M \times M$ matrix with diagonal elements $\left[\boldsymbol{\Sigma}_i[t]\right]_{mm}$ independently and uniformly distributed on $[1, 5]$.
- We defined the Average Square Error at time t ($ASE[t]$) to be the average value of $\sum_{i=1}^{N} \|\widehat{\boldsymbol{\theta}}_i[t] - \boldsymbol{\theta}\|_2/N$ over 500 implementations.

Result 1: Asymptotical performance of the proposed method

The error as a function of time for the proposed method with $(\epsilon_j = v_j / \max(deg(i))$, $v_1 = 0.9, v_2 = 0.5$ and $v_3 = 0.1$.



Diffusion Method (ATC)

At each time t, every agent received a scalar signal generated by the model $y_i[t] = \mathbf{h}_i[t]\boldsymbol{\theta} + w_i[t]$, and the estimates were updated by,

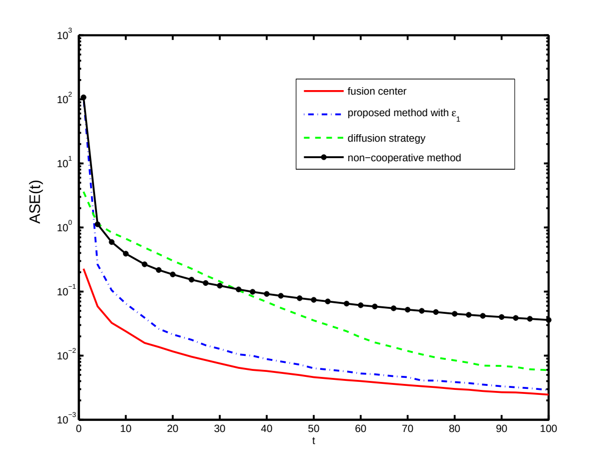
$$\breve{\boldsymbol{\theta}}_i[t] = \sum_{j=1}^{N} q_{i,j}\breve{\boldsymbol{\theta}}_i[t-1] + \mu \sum_{j=1}^{N} q_{i,j}\mathbf{h}_i[t]^\top \left[y_i[t] - \mathbf{h}_i[t]\breve{\boldsymbol{\theta}}_i[t-1]\right],$$

where the step size parameter $\mu = 0.01$.

Comparison between the ATC diffusion method and the proposed method

- Compared with our data model, the diffusion method takes M time periods to process the data that would take one time period to our method. In the simulation, when we plot the diffusion method, the time label is compressed by M times.
- By the diffusion method, the information exchanged between neighbors during these M time periods is M vectors of sizes $K \times 1$, while in the proposed method, it is a one $K \times 1$ vector and one $K \times K$ matrix.

Comparison (cont.)



Outline

Concluding remarks

- We presented an approach of distributed estimation by making use of the consensus method.
- We proved that the estimates held by the agents during the implementation of the proposed algorithm are unbiased.
- We also showed that the performance of every agent asymptotically reaches the performance of a fusion center.
- We demonstrated the performance of the method by computer simulations.
- Comparisons with the ATC method of Ali et al. were provided.