On the effect of informed nodes on learning over complex adaptive networks

Morteza Farhid[1,a], Mousa Shamsi[1,a], Mohammad Hossein Sedaaghi[1,a], Faramarz Alsharif[2,b], Bruno Senzio-Savino[2,c], and Mohammad Reza Alsharif[3,d]
[1]Faculty of Electrical Engineering, Sahand University of Technology, Tabriz, Iran
[2]Graduate School of Engineering and Science University of the Ryukyus, Okinawa, Japan
[3]Department of Information Engineering, Faculty of Engineering, University of the Ryukyus, Okinawa, Japan
E-mail: [a]{a_heydarpour, shamsi, sedaaghi}@sut.ac.ir, [b]faramarz_asharif@yahoo.com
[c]b.senzio@gmail.com, [d]asharif@ie.u-ryukyu.ac.jp

Abstract: In this paper, we investigate the impacts of number of informed nodes on the performance of a distributed estimation algorithm, namely adaptive-then-combine diffusion LMS, based on the data with the temporal and spatial independence assumptions. The study covers different network models, including the regular, the small-world and the random networks. We have two scenarios for our simulation. We change the fraction of nodes according to their links densities. The simulation results indicate that the larger proportion of the uninformed nodes (90% in first and up to 50% in second scenarios) in a network causes lower convergence besides improvement in the mean-square-error performance and that acquiring more information is not necessarily better.

Keywords: Adaptive networks, Diffusion LMS, Mean square deviation, Distributed estimation.

1. Introduction

Recently distributed signal processing has received much attention during the last decade [1-3]. The theory of decentralized systems is that the nodes organize themselves interacting locally and carry out the computations without the necessity of conveying the information to a Fusion Center (FC). Each node communicates with neighboring nodes -often located within a small range- to exchange their information and make decisions [4-5]. Distributed estimation is to estimate a vector of interest for each node, where the accuracy is improved by accessing to the measurements from a subset of its neighbors. This problem has been studied in the context of distributed control, tracking, in data fusion, and recently in wireless sensor networks [6-8]. In many applications, however, we need to perform estimation task in a constantly changing environment where the statistical information for the underlying processes of interest is not available. This motivates the development of distributed adaptive estimation schemes which are also known as adaptive networks. Adaptive networks consist of a collection of spatially distributed nodes that are linked together through a connection topology and that cooperate with each other through local interactions [9-10]. By means of cooperative processing in combination with adaptive filtering per node enables the entire network and also each individual node to track not only the variations of the environment but also the topology of the network [11]. Generally, according to the approach by which the nodes communicate with each other, distributed estimation schemes can be classified into incremental algorithms and diffusion algorithms (and also their probabilities). In the incremental mode, a cyclic path through the network is required, and nodes communicate with neighbors within this path. The incremental LMS, incremental RLS, incremental techniques based on the affine projection algorithm, parallel projections, and randomized incremental protocols are examples of incremental adaptive networks [9-10], [12-13]. The diffusion algorithms, however, allow each node to communicate with all of its neighbors as reflected by the network topology. Typical examples include diffusion LMS [11], diffusion RLS [14], and diffusion Kalman filtering [15]. As a cyclic pathway is no longer required, these algorithms are more preferable in practical engineering. The incremental-based networks present excellent estimation performance particularly in small size networks, while diffusion based networks are more robust to link and node failures.

In the previous works [10-15], the nodes in the network were assumed to be homogeneous in that all nodes had similar capabilities and were able to have continuous access to measurements. However, it is often observed in biological networks that the behavior of the network tends to be dictated more heavily by a small fraction of the agents, as happens with bees that it refers to as heterogeneous adaptive networks, where a fraction of the nodes are assumed to be informed while the remaining nodes are assumed to be uninformed. Informed nodes collect data and perform in-network processing, while uninformed nodes only participate in the processing tasks [16].

Yet, accompanying with the diffusion cooperative protocol over heterogeneous networks, a natural question has arisen: How does the number of informed nodes over complex networks affect the performance of distributed signal processing? Besides, answer to this question can help us in design the networks. In this work, we focus on the adaptive-then-combine (ATC) diffusion LMS, which has been proved to be superior to the other diffusion LMS algorithms [11]. The mean-square performances over different network models, including the regular, the small-world [17] and the random [18], are compared by numerical simulations. We also denote random quantities by boldface letters.

2. Distributed estimation based on diffusion LMS

We consider a connected network consisting of N nodes (Fig.1). Each node k collects scalar measurements $\mathbf{d}_k(i)$ and $1 \times M$ regression data vectors $\mathbf{u}_{k,i}$ over successive time instants with a positive definite covariance matrix, $R_{u,k} = E\boldsymbol{u}_{k,i}^{*}\boldsymbol{u}_{k,i}$. Two nodes are said to be neighbors if they can share information. The set of neighbors of node k including k itself is called the neighborhood of k and is denoted by N_k. The measurements across all nodes are assumed to be related to a set of unknown M × 1 vectors $\{w^o\}$via a linear regression model of the form [9]:

$$\boldsymbol{d}_k(i) = \boldsymbol{u}_{k,i} w^0 + \boldsymbol{v}_k(i), \quad k = 1,2,...N \tag{1}$$

where $\boldsymbol{v}_k(i)$ means measurement or model noise with variance $\sigma_{v,k}^2$ and assumed to be spatially and temporally white, i.e.,

$$E\boldsymbol{v}_k^*(i)\boldsymbol{v}_l(j) = \sigma_{v,k}^2 . \delta_{kl} . \delta_{ij} \tag{2}$$

in terms of the Kronecker delta function. The noise $\boldsymbol{v}_k(i)$ is also assumed to be independent of $\mathbf{u}_{l,j}$ for all l and j. All random processes are assumed to be zero mean and w^o denotes the parameter of interest for node k.

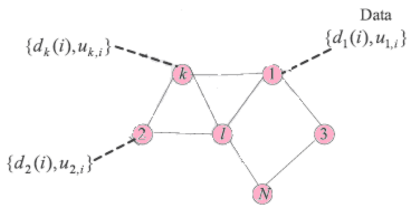

Fig. 1. Distributed network with N nodes. $\{d_k(i),u_{k,i}\}$ denotes the time realization for each node k.

For example, w^o can be the parameter vector of some underlying physical phenomenon, the location of a food source or a vector modeling different groupings of nodes. The nodes in the network are assumed to estimate the vectors $\{w^o\}$ by seeking the solution for the following minimization cost function:

$$\sum_{k=1}^{N} J_k(w) = \sum_{k=1}^{N} E\left|\boldsymbol{d}_k(i) - \boldsymbol{u}_{k,i} w\right|^2 \tag{3}$$

The objective is to estimate the vector of interest w^o from the data collected at N nodes spread in the network using the ATC diffusion LMS, which is originally proposed by [11]. It operates as follows. We assign an N × N matrix C with nonnegative entries $\{c_{l,k}\}$ that are real, non-negative constants satisfying:

$$C^T 1 = 1 \; and \; c_{l,k} = 0, \; if \; and \; only \; if, l \notin N_k \tag{4}$$

where 1 is a vector of size N with all entries equal to one. The entry $c_{l,k}$ denotes the weight on the link connecting node l to node k. The ATC algorithm consists of two steps namely adaptation and combination(Fig.2). In the adaptation step, each node k adaptively updates its estimate, denoted as $\phi_{k,i}$, with a steepest-descent implementation of the mean-square performance. Afterward, in the combination step, the node consults its peer nodes within its neighborhood and combines their estimates (denoted as $\left\{\phi_{l,i}; l \in N_k\right\}$, $\varphi_{k,i}$ where N_k is the set of nodes in the neighborhood of node k including itself) by a linear function to generate a new estimate $u_{k,i}$.

Mathematically, it is implemented as follows[11]:

$$\begin{cases} \varphi_{k,i} = \varphi_{k,i-1} + \mu_k u_{k,i}^{*}(d_k(i) - u_{k,i}\varphi_{k,i-1}) \\ \varphi_{k,i} = \sum_{l \in N_k} c_{l,k}\phi_{l,i} \end{cases} \tag{5a,5b}$$

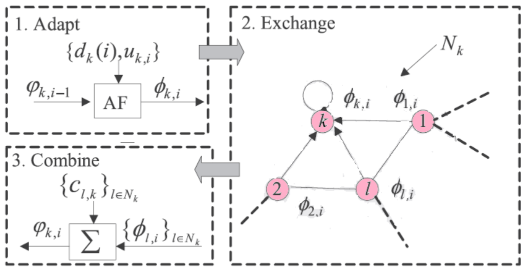

Fig. 2. ATC diffusion LMS algorithm.

μ_k is the positive step-size used by node k. Coefficients $c_{l,k}$ govern the node's cooperative rule, which are determined by the network topology. In regard to the combination protocols, several models, including the Metropolis rule, the relative degree, the Laplacian matrix, and adaptive combiners have been suggested. In this paper, we are interested in the following Metropolis rule as it is superior to the others [11]:

$$c_{l,k} = \begin{cases} 1/\max(n_k, n_l) & if \; l \in N_k \setminus k \\ 1 - \sum_{l \in N_k \setminus k} c_{l,k} & if \; l = k \\ 0 & if \; l \notin N_k \end{cases} \tag{6}$$

where $N_k \setminus k$ denotes the set of nodes in the neighborhood of node k excluding itself, n_k and n_l are the degrees for nodes k and l, respectively. To model heterogeneity over the network, like in [16], [19], we set $\mu_k = 0$ if node k is uninformed. In this model, uninformed nodes do not perform the adaptation step (5a) but continue to perform (5b).

The mean-square performance of the ATC algorithm was studied in detail by applying the energy conservation approach in [11] and the network mean-square-deviation (MSD) is used to assess how well the network estimates the weight vector, w°. The MSD is defined as follows:

$$MSD \triangleq \lim_{i \to \infty} \frac{1}{N} \sum_{k=1}^{N} E\left\| w^0 - \varphi_{k,i} \right\|^2 \tag{7}$$

The MSD expression for hetrogeneous adaptive networks is as[19]:

$$MSD \approx \frac{M\mu}{2N\eta} . \frac{\sum_{l \in N_I} \sigma_{v,l}^2 n_l^2}{\sum_{l \in N_I} n_l} + \frac{\mu^2 Tr(R_u) h(2/\sqrt{\eta})}{N\eta} \sum_{l \in N_I} \sigma_{v,l}^2 n_l^2 \tag{8}$$

We observe that the MSD in (45) depends on the network topology only through the average degree of the network η. Additionally, the MSD in (45) depends on the distribution of informed nodes through their degrees, n_l, and noise variances, $\sigma_{v,l}^2$. That is, the effect of different types of network models only depends on the degree distribution of the nodes.

We have performed a series of simulations to investigate the performance of the diffusion LMS algorithm over different kinds of complex networks from the viewpoint of mean-square errors.

3. Network models and data generation

Regular network [20]: It is generated from a regular nearest-neighbor network consisting of N nodes arranged in a ring, and each node has 2K nearest neighbors. The network corresponds to the original nearest-neighbor network when p = 0.

Small-world and random networks: In order to describe the transition from a regular lattice to a random graph, Watts and Strogatz proposed an interesting small-world network model, termed as WS small-world network. Links are then modified by rewiring one end to another node with a probability p while keeping another end unchanged. Nevertheless, no two nodes are allowed to be connected by more than one link. The network corresponds to the original nearest-neighbor network when almost like the ER random graph when p = 1.0. The degree distribution of the small-world network (0 < p < 1) follows a Poisson-like distribution. It peaks at an average value and decays exponentially. Such a network is also called homogenous network, as each node has nearly the same number of link connections.

For data generation model, we used these assumptions:

1) $\boldsymbol{u}_{k,i}$ is independent of $\boldsymbol{u}_{l,i}$ for $k \neq l$ (spatial independence).

2) For every k, the sequence $\{\boldsymbol{u}_{k,i}\}$ is independent over time (time independence).

3) The regressors $\{\boldsymbol{u}_{k,i}\}$ arise from a source with circular Gaussian distribution with covariance matrix $R_{u,k}$.

4. Simulation Results

We now apply the ATC diffusion LMS algorithm to estimate the unknown vector w^o from the data $\{d_k(i),u_{k,i}\}$ across all the N nodes in different kinds of complex heterogeneous adaptive networks. The small-world networks are generated by the WS algorithm with K = 2 and p = 0.1. The initial regular network (p = 0) and random network (p = 1.0) are also used for comparison. The measurements were generated according to model (1), and the regressors $\mathbf{u}_{k,i}$ were chosen Gaussian iid. The step size for informed nodes is set to 0.06 and size of the unknown vector w^0 is M=4 and variance of noise is set to 0.01. The results were averaged over 20 experiments. In Figs. 3-9, we show the effect of the number of informed nodes on the convergence factor and the MSD of the network. We increase the number of uninformed nodes, according to the highest density, i.e., from node 10% to node 90% in the regular model and up to 50% in other ones. The simulation results indicate that the larger the proportion of informed nodes in a network cause faster convergence rate at the cost of falling in the mean-square-error performance for ideal links and this is true for different number of neighbors.

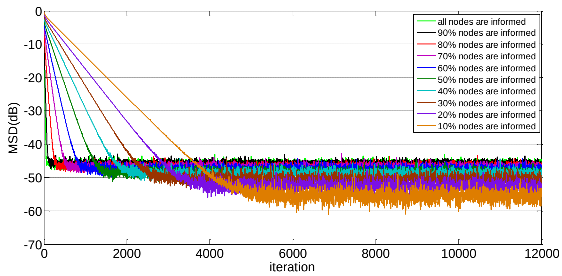

Fig.3.Transient network MSD in the regular model for different informed nodes.

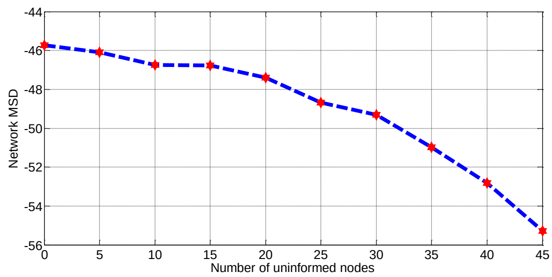

Fig.4.steady state of network MSD in the regular model for different informed nodes.

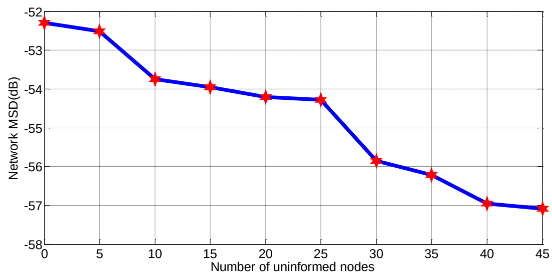

Fig.5.steady state of network MSD in the regular model with full connected for different informed nodes.

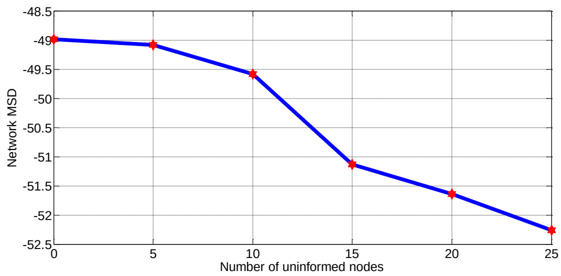

Fig.6. steady state of network MSD in the WS model with four neighbors for different informed nodes.

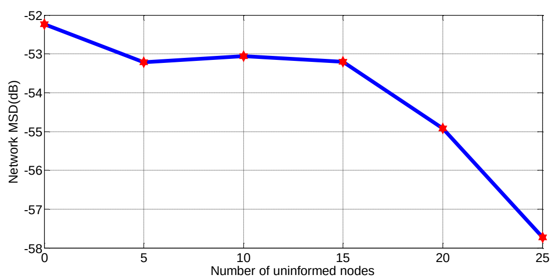

Fig.7.steady state of network MSD in the regular model with full connected for different informed nodes.

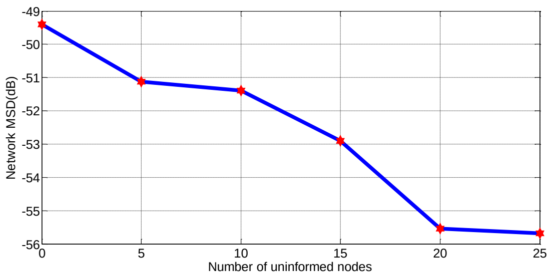

Fig.8. steady state of network MSD in the Random model with six neighbors for different informed nodes.

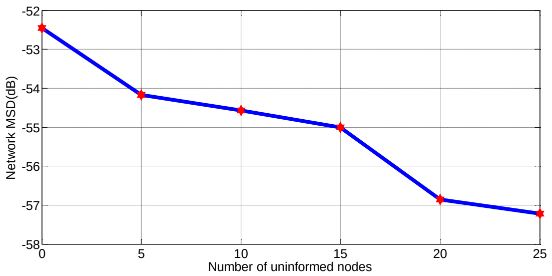

Fig.9. steady state of network MSD in the Random model with full connected for different informed nodes.

5. Conclusion

In this paper, we have investigated the performance of the ATC diffusion LMS algorithm over different heterogeneous network models from the viewpoints of mean-square performance. The results have shown that the larger the proportion of informed nodes in a network cause faster convergence rate at the cost of falling in the mean-square-error performance for ideal links. This fact helps us to design the networks and reach better steady state of MSD without increasing link density especially about the regular model.

References

[1] H. Goldingay, J. Mourika, " The effect of load on agent-based algorithms for distributed task allocation,", Inf. Sci., in press. doi:10.1016/j.ins, 2011.

[2] S. Ilarri, E. Mena, A. Illarramendi, " Using cooperative mobile agents to monitor distributed and dynamic environments," Inf. Sci. , vol. 178, no. 9, pp. 2105– 2127, 2008.

[3] N. Moghim, S.M. Safavi, M.R. Hashemi, " Performance evaluation of a new end-point admission control algorithm in NGN with improved network utilization," Int. J. Innov. Comput. Inf. Control, vol. 6, no. 7 , pp. 3067–3080, 2010.

[4] J. Liu, M. Chu, J.E. Reich, " Multitarget tracking in distributed sensor networks," IEEE Signal Process. Mag. , vol. 24, no 3, pp. 36–46, 2007.

[5] W. Sung, C. Chen, " Parallel data fusion for an industrial automatic monitoring system using radial basis function networks," Int. J. Innov. Comput. Inf. Control, vol 1, no. 6, pp. 2523–2536, 2010.

[6] A. S. Willsky, M. Bello, D. A Castanon, B.C Levy, G. Verghese, , " Combining and updating of local estimates and regional maps along sets of one-dimensional tracks," IEEE Trans. Automa.Control, vol. 27, no. 4, pp. 799–813, 1982.

[7] G. Mergen, L. Tong, , " Type based estimation over multiaccess channels," IEEE Trans. Signal Process, vol. 54, no. 2, pp. 613–626, 2006.

[8] A. Ribeiro, G.B. Giannakis, " Bandwidth-constrained distributed estimation for wireless sensor networks, Part I: Gaussian case," IEEE Trans. Signal Process,vol. 54, no. 3, pp. 1131–1143, 2006.

[9] C. G. Lopes, A. H. Sayed, " Distributed processing over adaptive networks," In: Proceeding of Adaptive Sensor Array Processing Workshop, MIT Lincoln Lab., Lexington, MA, 2006.

[10] C. G. Lopes, A. H. Sayed, " Incremental adaptive strategies over distributed networks," IEEE Trans. Signal Proces., vol. 55, no. 8, pp. 4064–4077, 2007.

[11] F. S. Cattivelli and A. H. Sayed, "Diffusion LMS strategies for distributed estimation," *IEEE Trans. on Signal Processing*, vol. 58, no. 3, pp. 1035–1048, Mar. 2010.

[12] A. H. Sayed, Lopes, C.G. , " Distributed recursive least-squares strategies over adaptive networks," In: Proceeding ofAsilomar Conference Signals, Systems, Computers, pp. 233–237, Monterey, CA , 2006.

[13] Li, L. Chambers, J.A, " A new incremental affine projection based adaptive learning scheme for distributed networks," Signal Process. Vol. 88, no. 10, pp. 2599–2603 , 2008.

[14] F.S. Cattivelli, C.G. Lopes, A. H. Sayed, " Diffusion recursive least-squares for distributed estimation over adaptive networks," IEEE Trans. Signal Process. Vol. 56,no. 5, pp.1865–1877, 2008.

[15] F.S. Cattivelli, C.G. Lopes, A. H. Sayed, " Diffusion strategies for distributed Kalman filtering: formulation and performance analysis," in: Proc. Workshop on Cognitive Inf. Process., Santorini, Greece, pp. 36–41, 2008.

[16] Sheng-Yuan Tu; A. H. Sayed, "On the Influence of Informed Agents on Learning and Adaptation Over Networks", Signal Processing, IEEE Transactions on, pp. 1339 – 1356, V. 61, no. 6, March15, 2013.

[17] M.E.J. Newman, D.J. Watts, " Renormalization group analysis of the small-world network model", Phys. Lett. A 263, pp. 341–346, 1999.

[18] P. Erdös, A. Rényi, " On the evolution of random graphs," Publ. Math. Inst. Hung. Acad. Sci. vol. 5, pp. 17–60, 1959.

[19] Sheng-Yuan Tu and Ali H. Sayed , " On the Effects of Topology and Node Distribution on Learning over Complex Adaptive Networks," IEEE, Asilomar, 2011.

[20] Ying Liu, Chunguang Li, Wallace K.S. Tang, Zhaoyang Zhang, "Distributed estimation over complex networks," Information Sciences, vol. 197, pp. 91–104, 2012.