Smart Structures Systems, *Vol. 19, No. 4 (2017) 383-391*
DOI: https://doi.org/10.12989/sss.2017.19.4.383

Distributed estimation over complex adaptive networks with noisy links

Morteza Farhid*, Mohammad H. Sedaaghi[a] and Mousa Shamsi[b]

Department of Electrical Engineering, Sahand university of technology, Tabriz, Iran

(Received June 6, 2016, Revised January 11, 2017, Accepted January 12, 2017)

Abstract. In this paper, we investigate the impacts of network topology on the performance of a distributed estimation algorithm, namely combine- then-adaptive (CTA) diffusion LMS, based on the data with or without the assumptions of temporal and spatial independence with noisy links. The study covers different network models, including the regular, small-world, random and scale-free whose the performance is analyzed according to the mean stability, mean-square errors, communication cost (link density) and robustness. Simulation results show that the noisy links do not cause divergence in the networks. Also, among the networks, the scale free network (heterogeneous) has the best performance in the steady state of the mean square deviation (MSD) while the regular is the worst case. The robustness of the networks against the issues like node failure and noisier node conditions is discussed as well as providing some guidelines on the design of a network in real condition such that the qualities of estimations are optimized.

Keywords: complex networks; adaptive networks; diffusion mode; distributed estimation; LMS algorithm

1. Introduction

Recently, distributed signal processing has received a great deal of attention due to a reduced computational requirements and power consumption (Goldingay and Mourika 2011, Ilarri *et al*. 2008, Moghim *et al*. 2010). The nodes organize themselves in distributed systems through local interactions and carry out the computations without the requirement to transfer the information to a Fusion Center (FC). Nodes, which are normally situated in a close-enough distance to each other, keep in touch in order to exchange their information and make decisions (Liu *et al*. 2007, Sung and Chen 2010). Distributed estimation is useful for the estimation of a desired vector for each node, leading to an improved accuracy by accessing to the measurements from a subset of its neighbors. This has been studied in the context of distributed control, tracking, data fusion, and recently in wireless sensor networks (Willsky *et al*. 1982, Mergen and Tong 2006).

There are, however, situations where the statistical information for the underlying processes of interest is not available. As a result, it is necessary to conduct estimation tasks in a constantly changing environment and the efforts to develop distributed adaptive estimation schemes (also known as adaptive networks) are then motivated. Adaptive networks consist of a collection of spatially distributed nodes that are linked together through a connection topology and cooperate with each other through local interactions (Ribeiro and Giannakis 2006, Lopes and Sayed 2006). When the cooperative processing is adopted in conjunction with adaptive filtering for each node, the entire network and also each individual node are enabled to track not only the variations of the environment but also the topology of the network (Lopes and Sayed 2007).

Distributed estimation schemes can be classified into incremental and diffusion algorithms (and also their probabilities) depending on the approach by which the nodes communicate with each other. In the incremental mode, a cyclic path through the network is required, and the 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 (Lopes and Sayed 2006, 2007, Sayed and Lopes 2006, Li and Chambers 2008). In contrast, nodes are allowed to communicate with all of their neighbors in the diffusion algorithms as reflected by the network topology. Typical examples include diffusion LMS (Cattivelli and Sayed 2010), diffusion RLS (Cattivelli *et al*. 2008) and diffusion Kalman filtering (Cattivelli *et al*. 2008). Since a cyclic pathway is no longer required, these algorithms are preferred in practical engineering. We should note that incremental-based networks are particularly accurate in small size networks, while diffusion based networks are more robust to link and node failures.

There are numerous evidences of complex networks in modern science, including the study of biological networks, power grids, macro-economies and inference over graphs. In many complex systems, especially those encountered in nature, it is common for emergent behavior to arise from the interaction among individual agents, as happens with fish schooling or bird flight formations. While each

*Corresponding author, Ph.D. Student
E-mail: m_farhid@sut.ac.ir
[a]Ph.D.
E-mail: sedaaghi@sut.ac.ir
[b]Ph.D.
E-mail: shamsi@sut.ac.ir

http://www.techno-press.com/journals/sss&subpage=7
ISSN: 1738-1584 (Print), 1738-1991 (Online)

individual agent in these biological networks is not capable of featuring complex behavior, it is the combined coordination among multiple agents that leads to the manifestation of sophisticated behavior at the network level. Research efforts to decipher the intricacies of such complex networks have been progressing almost independently across several disciplines, including signal processing, machine learning, optimization, control, statistics, physics, biology, economics, computer science, and the social sciences. Yet, in all of the mentioned discipline above, the highest interest is towards performing inference and learning over graphs (Strogatz 2001, Wang 2002). In the realm of signal processing, these applications prompt the need to study and develop decentralized strategies for information processing that are able to endow networks with real-time adaptation and learning abilities. In implemented adaptive-then-combine (ATC) diffusion least mean square (LMS) over complex networks (Ying *et al*. 2012), the data are assumed to be exchanged among neighboring nodes without distortion. This assumption, however, may not be true in practice due to the noisy link. The distributed estimation problem with noisy links has also been considered in the context of consensus-type and diffusion algorithms (Kar and Moura 2009, Touri and Nedic 2009, Khalili *et al*. 2012). However, to the best of our knowledge, the same issue has not yet been studied in the context of distributed estimation over complex networks. Questions are remained to be addressed in real conditions: i) the suitable topology and ii) the type of the network to be designed (homogeneous or heterogeneous).To answer these questions, a methodical study is carried out in this paper to analyze the impacts of the network topology on the performance of distributed estimation with noisy links. We focus on the combine-then-adapt (CTA) diffusion LMS and mathematically analyze the mean stability of this algorithm, while the mean-square performances over different network models, including the regular, small-world (Newman and Watts 1999), random (Erdös and Rényi 1959) and the scale-free (Barabási and Albert 1999) are compared by numerical simulations with noisy links. The rest of the paper is organized as follows. In Section 2, we briefly revisit the CTA diffusion LMS strategy. Its performances in terms of mean stability and mean-square errors are then discussed in Section 3. In Section 4, the performances of the CTA diffusion LMS over different network models are presented and the results are discussed in detail. Robustness of the networks from node failure, noisier nodes and link density effect are simulated in section 5. Finally, conclusions are drawn in Section 6.

Notation: We adopt boldface letters for random quantities. The symbol * denotes conjugation for scalars and Hermitian transpose for matrices. The notation col{} denotes a column vector (or matrix)with the specified entries stacked on top of each other. We also denote by A⊗B the Kronecker product of two matrices A and B.

2. Distributed adaptive diffusion estimation

We consider a connected network consisting of N={1, 2, …, N} nodes (see Fig. 1). All nodes are assumed to measure data that satisfy a linear regression model of the form

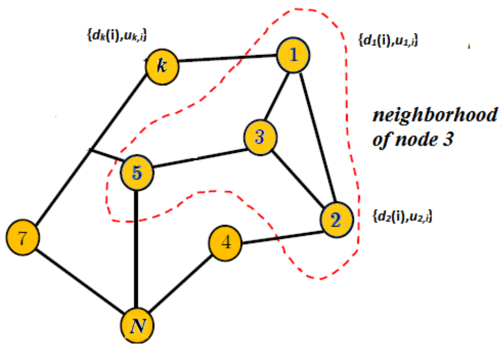


Fig. 1 Distributed network with N nodes. $\{\boldsymbol{d}_k(\mathrm{i}),\boldsymbol{u}_{k,\mathrm{i}}\}$ denotes the time realization for each node k

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

where w^{o} is a deterministic but unknown M×1 vector, $\boldsymbol{d}_k$(i) is a random measurement datum at time i, $\boldsymbol{u}_{k,i}$ is a random 1×M regression vector at time i, and $\boldsymbol{v}_k$(i) is a random noise signal also at time i. The statistical assumptions of the data $\{\boldsymbol{u}_{k,i}, \boldsymbol{v}_k(\mathrm{i})\}$ are as follows:

The regression data $\boldsymbol{u}_{k,i}$ are temporally white and spatially independent random variables with zero mean and uniform covariance matrix $R_{u,k} = E\boldsymbol{u}_{k,i}^{*}\boldsymbol{u}_{k,i}$. The noise signals $\boldsymbol{v}_k$(i) are temporally white and spatially independent random variables with zero mean and variances $\sigma_{v,k}^2$. The regressors $\boldsymbol{u}_{k,i}$ and noise signals $\boldsymbol{v}_k$(i) are mutually-independent for all k and l, i and j.

Two nodes are said to be neighbors if they can contribute to information. The set of neighbors of node k including k itself is called the neighborhood of k and is denoted by N_k.

For example, w^{o} can be the parameter vector of some underlying physical phenomenon, the location of a food source(Tu and Sayed 2011) 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

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

Obviously, the objective of the network is not to use $w^0 = (\sum_{k=1}^{N} R_{u,k})^{-1}(\sum_{k=1}^{N} r_{du,k})$, where $R_{u,k} = \mathrm{E}\boldsymbol{u}_{k,i}^{*}\boldsymbol{u}_{k,i} > 0$ and $r_{du,k} = \mathrm{E}\boldsymbol{d}_k(i)\boldsymbol{u}_{k,i}^{*}$, to determine w^{o} since this would require that each node has access to the second-order moments from across the entire network(Lopes and Sayed 2006). Instead, the nodes in the network would like to rely solely on data realizations that are available to them locally in order to estimate. So the aim is to estimate the vector of interest w^{o} from the data collected at N nodes spread in the network using the CTA diffusion LMS, which is originally proposed by (Cattivelli

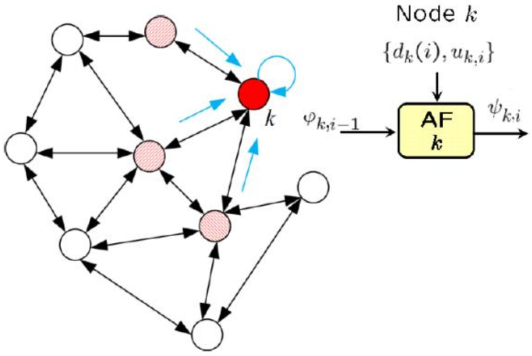


Fig. 2 CTA diffusion algorithm LMS

and Sayed 2010). It operates as follows. We assign an $N\times N$ matrix C with nonnegative entries $\{c_{l,k}\}$ that are real, non-negative constants satisfying

$$C^T 1 = 1 \text{ and } c_{l,k} = 0, \text{ if and only if } , l \notin N_k \qquad (4)$$

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. In other words, $\{c_{l,k}\}$ is zero if nodes are not connected and the entries on each row of C add up to one.

Diffusion LMS algorithms use the data to find an estimate for the unknown vector. There are two different strategies for this estimation. The first strategy collects estimates from its neighbors from the previous iteration and combines them using some convex combiner method. Then the combined estimate is used to calculate a new estimate using the available data $\{\boldsymbol{d}_k(\mathrm{i}), \boldsymbol{u}_{k,i}\}$. Mathematically, it is implemented as follows (Khalili *et al*. 2012)

$$\begin{aligned} \varphi_{k,i-1} &= c_{kk}(i)\psi_{k,i-1} + \sum_{l\in N_k} c_{l,k}(\psi_{l,i-1} + \boldsymbol{q}_{l,k,i-1}) \\ \psi_{k,i} &= \boldsymbol{\varphi}_{k,i-1} + \mu_k \boldsymbol{u}_{k,i}^{*}(\boldsymbol{d}_k(i) - \boldsymbol{u}_{k,i}\boldsymbol{\varphi}_{k,i-1}) \end{aligned} \qquad (5)$$

where $\boldsymbol{\varphi}_{k,i-1}$ is the combined estimate and $\psi_{k,i}$ is the estimate for node k at iteration i and $\boldsymbol{q}_{l,k,i-1}$ are time-realizations of zero-mean wide-sense stationary random process with covariance matrices $Q_{l,k} = \mathrm{E}\boldsymbol{q}_{l,k}\boldsymbol{q}_{l,k}^{*}$. This is known as Combine-then-Adapt (CTA) algorithm with noisy links. The CTA algorithm consists of two steps namely combination and adaptation (see Fig. 2). μ_k is the positive step-size used by node k. Coefficients $c_{l,k}$ govern the node's

Table 1 Possible combination rules for $c_{l,k}$. n_i denotes the degree of node i

Combination rule	definition
Constant	$c_{l,k} = 1/n_k$
Laplacian	$c_{l,k} = 1/n_{\max}$
Metropolis	$c_{l,k} = 1/\max(n_k, n_l)$
Maximum degree	$c_{l,k} = 1/\mathrm{N}$
Relative degree	$c_{l,k} = n_l / (\sum_{m\in N_k} n_m)$

cooperative rule, which are determined by the network topology. With regard to the combination protocols, several models, including the Metropolis rule, the relative degree, the Laplacian matrix, and adaptive combiners have been suggested.

Table 1 depicts several ways to select the combination weights, motivated by graph theoretical quantities and common approaches in the literature.

In this paper, we use mostly Metropolis rule as it is superior to the others (Ying *et al*. 2012). Also we further assume that when a link between two nodes is available, it is noisy.

3. Performance analysis

The mean stability analysis aims to find out the sufficient conditions such that the local estimate at each node converges in the mean to the unknown parameter w^o.

Let the error vector for any node k be $\tilde{\psi}_{k,i} = w^o - \psi_{k,i}$. We collect all weight error vectors and step-sizes across the network into a block vector and block matrix: $\tilde{\psi}_i$ = col{ $\tilde{\psi}_{k,i}$ } and M= diag$\{\mu_k I_M\}$, and introduce the extended combination matrix $C=C\otimes I_M$ where the symbol $\otimes$ denotes the Kronecker product operation of two matrices. Then, starting from (5) and using model (1), we can verify that the global error vector evolves according to the relation

$$\tilde{\psi}_i = C\tilde{\psi}_{i-1} - q_{i-1} - MU_i^{*}(\mathrm{U}_i C\tilde{\psi}_{i-1} + v_i) + MU_i^{*}\,\mathrm{U}_i\, q_{i-1} \qquad (6)$$

Assuming temporal and spatial independence of the regressors and taking expectations of both sides leads to

$$\mathrm{E}\tilde{\psi}_i = C(I_{NM} - MR_u)\mathrm{E}\tilde{\psi}_{i-1} \qquad (7)$$

where R_u=diag$\{R_{u,1}, \ldots, R_{u,N}\}$ is block diagonal. Therefore, the mean evolution of the global weight error vector depends on the data moment $M\,\mathrm{E}\,\boldsymbol{u}_{k,i}^{*}\boldsymbol{u}_{k,i}$ and on the mean topology matrix EC_i, with changing topology of the networks. Henceforth, for stability in the mean we must have that

$$|\rho(C(I_{NM} - MR_u))| < 1 \qquad (8)$$

In other words, the spectrum of $(I_{NM}$-$MR_u)C$ must be strictly inside the unit disc. In the adaptive network case, even convergence in the mean will effectively depend on space-time data statistics and network topology (represented by C). For simplicity, assume that D=μI_{NM} so that $I_{NM} - MR_u$ is Hermitian (Cattivelli and Sayed 2010). Using matrix 2-norms we have

$$\|\mathrm{C}(I_{NM} - MR_u)\|_2 \le \|C\|_2 . \|(I_{NM} - MR_u)\|_2 \qquad (9)$$

That is, $\sigma_{\max}(BC) \le \sigma_{\max}(B)$. $\sigma_{\max}(C)$ (Cattivelli and Sayed 2010), where $\sigma_{\max}$ is the maximum singular value of the corresponding matrix. For combiners that render stochastic and symmetric matrices C, the matrix G will also be symmetric and stochastic so that $\sigma_{\max}(G)$=1. But since $\sigma_{\max}(B)$=$|\lambda_{\max}(B)|$ and $|\lambda_{\max}(BC)| \le \sigma_{\max}(BC)$, we conclude that

$$|\rho(C(I_{NM} - MR_u))| < |\rho((I_{NM} - MR_u))| \qquad (10)$$

That is, the spectral radius of BG is generally smaller than the spectral radius of B. Hence, cooperation under the diffusion protocol (3) has a stabilizing effect on the network.

The mean-square performance of the CTA algorithm was studied in detail by applying the energy conservation approach in (Cattivelli and Sayed 2010) and the network mean-square-deviation (MSD) and excess-mean-square-error (EMSE) are used to assess how well the network estimates the weight vector, w^o. The MSD and EMSE are defined as follows

$$MSD \triangleq \lim_{i\to\infty} \frac{1}{N}\sum_{k=1}^{N} E\left\| w^0 - \varphi_{k,i} \right\|^2 \tag{11}$$

$$EMSE \triangleq \lim_{i\to\infty} \frac{1}{N}\sum_{k=1}^{N} E\left\| \boldsymbol{u}_{k,i}(w^0 - \varphi_{k,i}) \right\|^2 \tag{12}$$

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.

4. Network models

We overview some common topologies assumed in the study of graphs. The following models apply to both undirected and directed graphs.

Regular network which 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. A ring (special case) is a one-dimensional grid where the vertices are spatially distributed forming a circle. The ring is a 2-regular graph, since every node has exactly two neighbors (Newman and Watts 1999).

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. A common feature of the ER random graph and the WS small-world models is that the connectivity distribution of the network is homogenous, with peak at an average value and decay exponentially. Such networks are called homogenous networks.

To explain the origin of power-law degree distribution, another network model (BA) is proposed by (Barabási and Albert 1999). They argued that many existing models fail to take into account two important attributes of most real networks. First, real networks are open and they are dynamically formed by continuous addition of new nodes to the network; but the other models are static in the sense that although edges can be added or rearranged, the number of nodes is fixed throughout the forming process. Second, both the random graph and small-world models assume uniform probabilities when creating new edges, but this is not realistic either. The BA model suggests that two main ingredients of self-organization of a network in a scale-free structure are growth and preferential attachment.

It starts with a small network composing of $N(0)$ nodes and $L(0)$ links. A new node is then added in each step and linked with m existing nodes using the standard preferential attachment mechanism until a network of N nodes is obtained.

After t time steps, the algorithm results in a network with $N=N(0)+t$ nodes and $L=L(0)+mt$ links. Studies in scale free networks have shown that the nodes' degrees follow a power-law distribution and thus the network is free of characteristic scale (Barabási and Albert 1999). In these graphs, some vertices are highly connected but most vertices have a low number of connections. Consequently, it is referred as heterogeneous network, because most nodes have very few link connections except a few high-degree nodes.

The last topology that we mention here is a lattice where the vertices are spatially distributed according to a two dimensional grid. The number of neighbors for an internal vertex is 4, whereas for an external vertex is 2.

5. Simulation results

We now apply the CTA diffusion LMS algorithm to estimate the unknown vector w^o from the data $\{\boldsymbol{d}_k(\mathrm{i}),\boldsymbol{u}_{k,i}\}$ across all the N nodes in different kinds of networks. The small-world networks are generated by the WS algorithm with K=2 and p=0.1. In this way, a total of L=2000 links are generated. To make a fair comparison with the work of (Ying *et al*. 2012), we used regular network (p=0) and random network (p=1.0) and set $N(0)$=5, $L(0)$=5, and m=2 for the BA scale-free network. After t=995 steps, we have a total of 1000 nodes and 1995 links. Since removing an edge may reduce the spectral radius of the Laplacian, which in turn improves the asymptotic speed of convergence of the matrix GB, the same values of N and L must be used for each model for an exact comparison. Therefore, five more links are randomly added into the BA model so that the total number of links also becomes 2000. For such a large network, we assume that this minor change will not affect its scale-free property.

• Independent regressors

In this case, we assume that the regressors' data arise from an independent Gaussian where their eigenvalue spread is ρ=5. Although the analysis relied on the independence assumptions, simulations presented in this subsection were carried out using regressors with shift structure to better represent realistic scenarios.

• Regressors with shift structure

For this definition of regressors, we use simulations using regressor input with shift structure. The desired data are generated according to the model given in (1), and the unknown vector w^o is set to $[1,1,\ldots,1]^T / \sqrt{M}$. In the same example as in (Chunguang *et al*. 2012) for simulations to for facilitate the comparison. We assume that the non-Gaussian regressors are uniformly distributed and are generated at each node k according to the recursion

$$u_k(i) = a_k . u_k(i-1) + b_k . n_k(i) \tag{13}$$

Eq. (13) describes a first-order autoregressive process with a pole at α_k that $a_{\mathbf{k}} \in [\mathbf{0}, .5)$ is the correlation index and $n_k(i)$ is a spatially independent white Gaussian process with unit variance and $b_k = \sqrt{\sigma_{u,k}^2 (1-a_k^2)}$. The regressor power profile is considered as $\sigma_{u,k}^2 \in (0,.5]$. The resulting regressors have Toeplitz covariance with correlation sequence $r_k(i) = \sigma_{u,k}^2 (a_k)^{(i)}, i = 0,1,2,\ldots,M-1$. These parameters are chosen randomly and taken the same as in (Chunguang *et al*. 2012) for comparison purpose.

The mean stability based on the criterion of spectral radius of the matrix GB, as given in (11), is firstly investigated. In the simulation, an i.i.d. Gaussian noise with variance $\sigma_{v,k}^2 = 1\times 10^{-3}$ is considered and a small step-size μ_k=0.03 is adopted. The length of the regressor is set as M=3, and thus there are total 3000 eigenvalues of the matrix *GB*, denoted as $\rho(GB)$. Fig. 3 depicts these eigenvalues for different networks, where the eigenvalues of the matrix B, denoted as ρ(B), correspond to the case without cooperation. The results are averaged over 40 times of independent experiments. As illustrated in Fig. 3, we have the spectral radii (maximum eigenvalue norm) satisfy $q_{\mathrm{Ran}}<q_{\mathrm{WS}}<q_{\mathrm{BA}}<q_{\mathrm{RG}}<q_{\mathrm{NC}}<1$, (while q_{RG}, q_{WS}, q_{Ran} and q_{BA} denote diffusion cooperative matrices *GB* for the regular, the WS small-world, the random and the BA scale-free networks, respectively) and hence the convergence of the estimate in the mean sense can be assured. Moreover, it is noticed that the cooperation reduces the eigenvalues as compared with the non-cooperative scheme, confirming our analysis given in Section 3. We should also mention the inequality applied for the spectral radii can vary depending on the combination coefficient, as happens for curves in the right panel of Fig. 3. As Fig. 3 shows, the cooperation significantly decreases the eigenmodes and thus yields faster convergence as compared with the non-cooperative scheme. Generally this is the largest eigenmodes that determine the speed of convergence and the resulting speed in Regular networks, compared to the other ones, is higher.

Fig. 4 plots the transient MSD of the entire network in dB for different network models. From Fig. 4, we notice that the regular network has the worst MSD performance. However, by rewiring a certain number of links, the accuracy of the estimation is significantly improved with the WS small-world network (p=0.1). The results are even better when p=1.0, corresponding to the random network. The performance of BA scale-free network lies somewhere between the regular and the small-world. It is also worth pointing out that the results in Figs. 3 and 4(a) are consistent. With the decrease of q(GB), the convergence rate is increased and the accuracy of the estimate is improved. Based on the complex network theory, all these networks have shorter average path length as compared with the regular networks (Wang 2002). This suggests that the average path length is one of the key factors affecting the performance of in-network distributed estimation. In general, the shorter the average path length, the better the

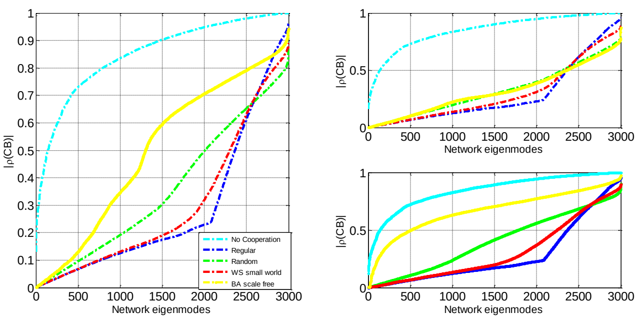
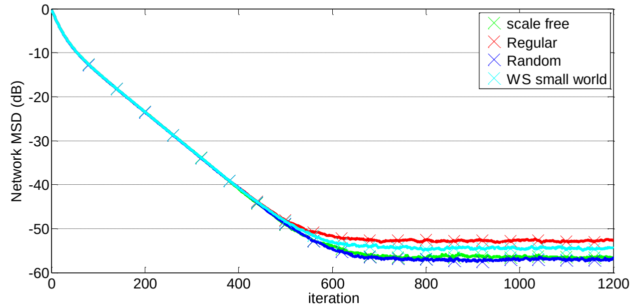
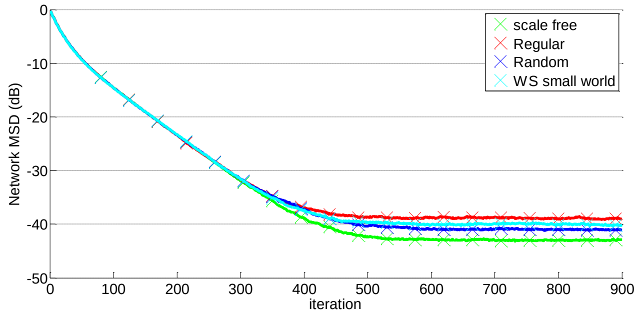


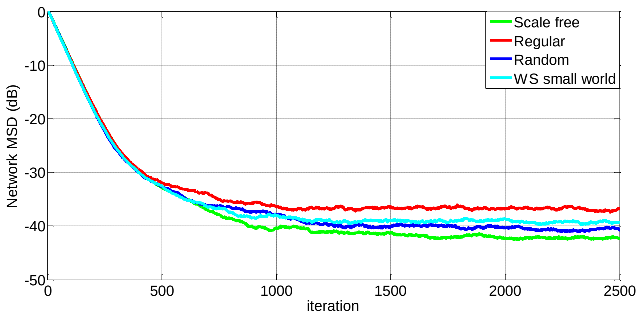
Fig. 3 The eigenvalues of different network models for three different combination rule (Metropolis (left), uniform (right, top) and laplacian (right, bottom)) with defined the network



(a) Ideal links



(b) Noisy links (Q=10^{-3}I), temporal and spatial independence



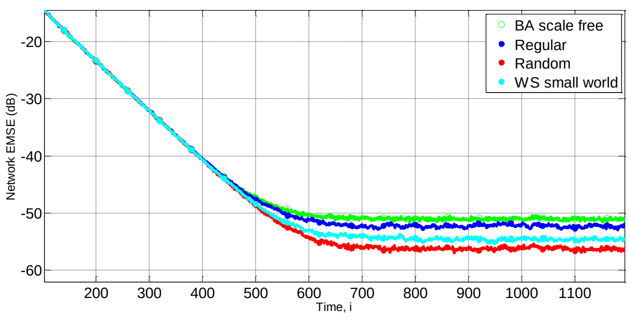
(c) Noisy links (Q=10^{-3}I), temporal and spatial dependence

Fig. 4 The Learning behaviors of the CTA diffusion LMS algorithm under different conditions performance of the distributed estimation.

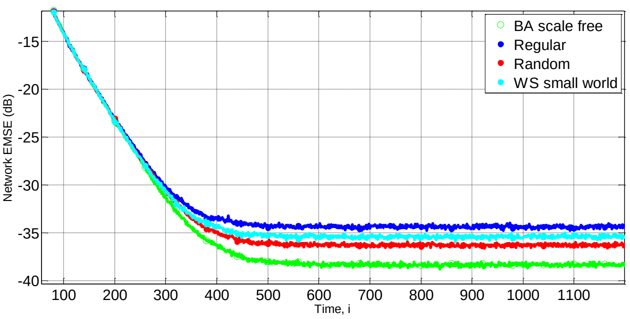
In Fig. 4, we have shown the MSD and EMSE learning curves for two different conditions including (1) networks with noisy links and (2) networks with ideal links. We assume that channel noise terms $\boldsymbol{q}_{k,i-1}$ are temporally and spatially i.i.d. with circular white Gaussian distribution.

As shown in Figs. 4-5, the MSD performances under the independent and dependent conditions are quite similar. In all of the following simulations, the step-size is set to μ_k=0.03 and only the case with temporal and spatial independence is considered. However, similar conclusions can be extended to other cases such as temporally and spatially dependent data in noisy condition of channels. As it is understood from Figs. 4-5, the scale free network, compared to the regular ones, shows a better MSD value by about 20 percent in real condition (noisy channels) and this is a motivation to design networks with heterogeneous link density.

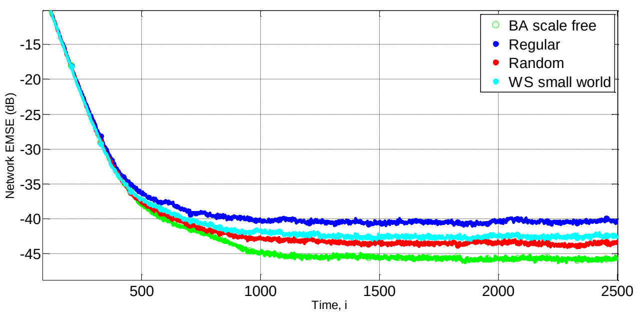
We also investigate the impact of the step-size μ_k to the performance of the estimation for different networks and



(a) Ideal links



(b) Noisy links(Q=10^{-3}I), temporal and spatial independence



(c) Noisy links (Q=10^{-3}I), temporal and spatial dependence

Fig. 5 The Learning behaviors of the CTA diffusion LMS algorithm under different conditions

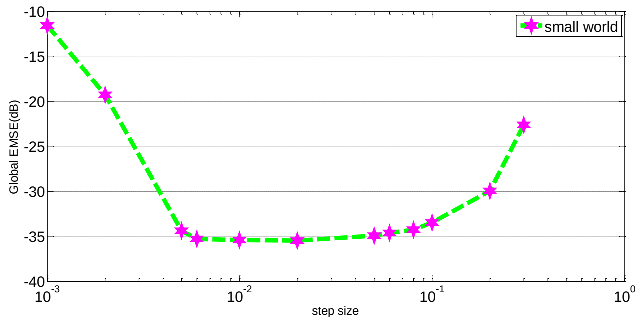
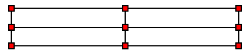
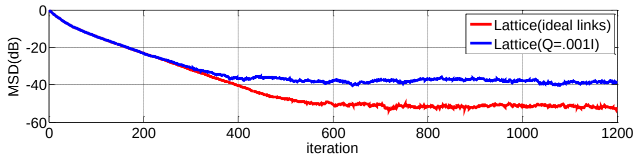
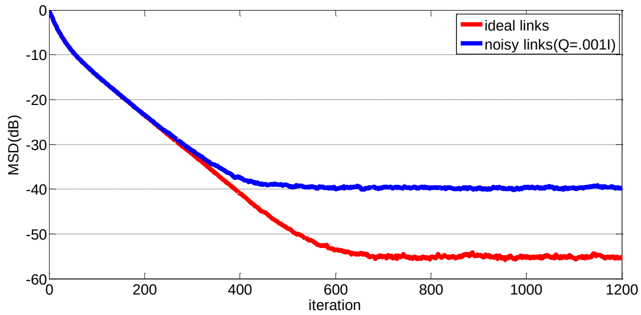


Fig. 6 The EMSE performance against the change of the step-size for a small-world network





(a) one lattice structure



(b) defined lattice with 1000 nodes

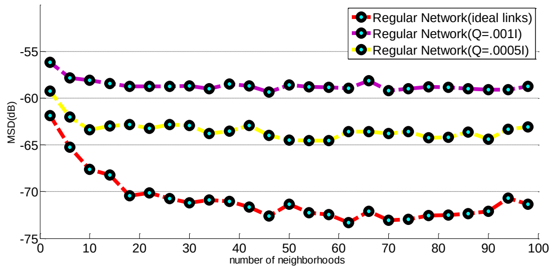
Fig. 7 Learning behaviors of the CTA diffusion LMS algorithm over lattice network

only small world network is presented in Fig. 6. As shown in Fig. 6 the EMSE performance is not monotonically increasing (ideal links) is in agreement with the conclusion obtained in (Khalili *et al*. 2012). As it is shown in this figure, the steady state performance of the network in this condition depends on the step sizes and there are some parameter values for which, the non-cooperative network works better.

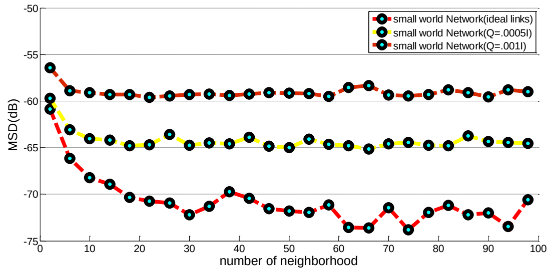
As explained in the network models for lattice structure, we simulated the estimation performance for one lattice structure (Fig. 7, left) and also with predefined the network for faire comparison. As illustrated here, this kind of networks' performance is similar with scale free network and has better performance in noisy channel in comparison to other ones.

5.1 Links density

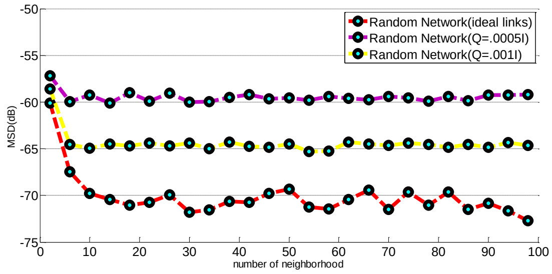
To show the relation between number of neighbors with steady state of MSD (estimation result), we defined three networks with 100 nodes and perform estimation with



(a) regular network



(b) small world network



(c) random network

Fig. 8 steady state of MSD with links density varation

increasing number of neighborhood until it became fully connected. As depicted in Figs. 8, by increasing the number of neighbors (link density) the estimation of the networks does not improve monotonically. It means that about 30% of a fully connected version of the networks we can reach the optimal design from the estimation performance point of view and implementing of networks budget (power consumption and communication burden). However, with increasing the noise in links, increasing this value (link density) may not improve the result of estimation performance. As it is demonstrated in this simulation results, in all of the networks with increasing link density up to maximum number of neighbors (full connected), we obtain improvements in steady state up to 12dB in ideal links which decreases by the presence of noisy links.

5.2 Effect of noisier nodes

More specifically, we consider the estimation problem in an inhomogeneous environment where some of the nodes make unreliable observations (noisy nodes). To show the effects of noisy nodes we consider a distributed network with 100 nodes. We also assume that up to 90% of nodes may become noisier with variance of observation noise belonging to (0, 2) based on maximum numbers of links. As

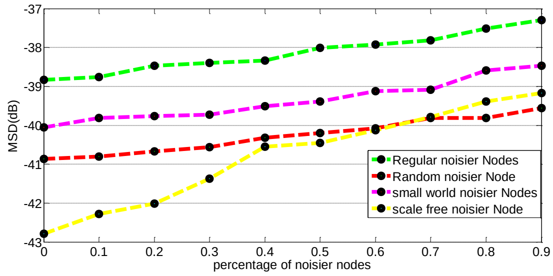


Fig. 9 steady state of MSD with noisier nodes

shown in Fig. 9, in this condition, noisy nodes do not deteriorate considerably the performance of the CTA algorithm (except scale free) and this proves one advantage for these networks. Although this degradation can be dependent on the values of noisier nodes, the behavior is all the same as depicted in this figure.

5.3 Node failure

In a practical sensor network, a sensor may be out of power, damaged or attacked, making its measurement unreliable. If this happens, the sensor will only observe the pure noise and certainly degrade the estimation performance. We refer to this phenomenon as node failure. Once a node fails to work, it is assumed to be turned silent and no information is exchanged with the others. This is simulated by removing this node together with its links from the network. The node removal is studied here and we name it intentional removal where the nodes of highest-degree are removed.

The steady-state network MSD is used as the measure to evaluate the robustness of different networks. In Fig. 10 we plot MSD variation by increasing the number of intentional removals with noisy links ($Q=10^{-3}I$). From Fig. 10, we find that the homogenous networks, i.e. the random and the small-world, show similar tendencies for intentional removal while the robustness of the small-world network is slightly better.

5.4 Link-rewiring experiment

Our simulations are performed by starting with a regular network of N=1000 nodes, where each node is coupled to its 2K=4 nearest neighbors. The rewiring probability p is then varied from 0 (regular) to 1 (random) with ideal and noisy links. From Fig. 11, it can be observed that in the

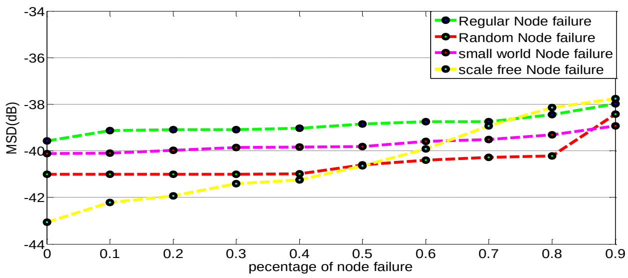


Fig. 10 steady state of MSD with node removal condition

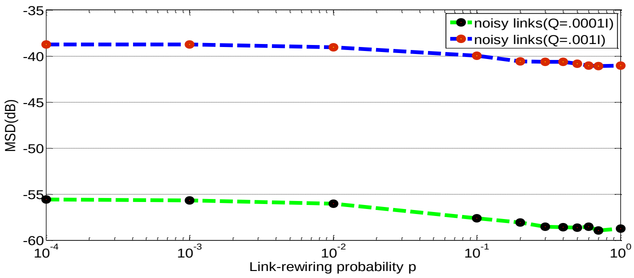


Fig. 11 steady state of MSD with link rewiring probability

noisy links case, the random networks (p=1) work well in homogeneous networks.

6. Conclusions

The importance of studying complex networks in the signal processing and learning arises from the fact that cooperation among nodes in real networks seems to improve certain performance inference value like estimation and detection. In this study, a numerical simulation procedure for distributed adaptive estimation over complex networks has been proposed. We have demonstrated that the presence of noisy links do not affect the convergence of the introduced networks. While the scale free networks have better performance in real working conditions (noisy links) in terms of steady state value of MSD, regular networks are the worst, both in ideal and in noisy links. We have shown that the trend of MSD value versus the step size is not monotically increasing like in ideal links networks. This helps in obtaining the value of MSD through properly selection of the step sizes. It has also proposed to use lattice structure networks which have similar MSD behavior to that of the scale free networks. In the view of design from aspects of efficient consumption of sources (power and bandwidth), it has also shown that, unlike ideal links where about 30% of neighbors connectivity is equal to full connection of nodes, it can be reduced about 10% giving the same estimation performance in the presence of noisy links.

Also given are the effect of noisier nodes, node failure according to their link densities and link rewiring probability effect on the overall estimation performance.

References

Barabási, A.L. and Albert, R. (1999), “Emergence of scaling in random networks”, *Science*, **286**(15), 509-512.

Cattivelli, F.S. and Sayed, A.H. (2010), “Diffusion LMS strategies for distributed estimation”, *IEEE Trans. Sign. Proc.*, **58**(3), 1035-1048.

Cattivelli, F.S., Lopes, C.G. and Sayed, A.H. (2008), “Diffusion recursive least-squares for distributed estimation over adaptive networks”, *IEEE Trans. Sign. Proc.*, **56**(5), 1865-1877.

Cattivelli, F.S., Lopes, C.G. and Sayed, A.H. (2008), “Diffusion strategies for distributed Kalman filtering: formulation and performance analysis”, *Proceeding of the Workshop on Cognitive Infrastructure Process*, Santorini, Greece.

Erdös, P. and Rényi, A. (1959), “On the evolution of random graphs”, *Pub. Math. Inst. Hung. Acad. Sci*., **5**(1), 17-60.

Goldingay, H. and Mourika, J. (2011), “The effect of load on agent-based algorithms for distributed task allocation”, *Inform. Sci*., **222**, 66-80.

Horn, R.A. and Johnson, C.R. (1990), “Matrix analysis”, *Cambridge*, England.

Ilarri, S., Mena, E. and Illarramendi, A. (2008), “Using cooperative mobile agents to monitor distributed and dynamic environments”, *Inform. Sci.*, **178**(9), 2105-2127.

Kar, S. and Moura, J.M.F. (2009), “Distributed consensus algorithms in sensor networks with imperfect communication: Link failures and channel noise”, *IEEE Trans. Sign. Proc.*, **57**(1), 355-369.

Khalili, A., Tinati, M.A., Rastegarnia, A. and Chambers, J.A. (2012), “Steady-state analysis of diffusion LMS adaptive networks with noisy links”, *IEEE Trans. Sign. Proc.*, **60**(2), 974-979.

Khalili, A., Tinati, M.A., Rastegarnia, A. and Chambers, J.A. (2012), “Steady-state analysis of diffusion LMS adaptive networks with noisy links”, *IEEE Trans. Sign. Proc.*, **60**(2), 974-979.

Khalili, A., Tinati, M.A., Rastegarnia, A. and Chambers, J.A. (2012), “Transient analysis of diffusion least-mean squares adaptive networks with noisy channels”, *Wiley Int. J. Adapt. Control Sign. Proc*., **26**(2), 171-180.

Li, L and Chambers, J.A. (2008), “A new incremental affine projection based adaptive learning scheme for distributed networks”, *Sign. Proc*., **88**(10), 2599-2603.

Liu, J., Chu, M. and Reich, J.E. (2007), “Multitarget tracking in distributed sensor networks”, *IEEE Sign. Proc. Mag.*, **24**(3), 36-46.

Lopes, C.G. and Sayed, A.H. (2006), “Distributed processing over adaptive networks”, *Proceeding of Adaptive Sensor Array Processing Workshop, MIT Lincoln Lab*., Lexington, MA.

Lopes, C.G. and Sayed, A.H. (2007), “Incremental adaptive strategies over distributed networks”, *IEEE Trans. Sign. Proc*., **55**(8), 4064-4077.

Mergen, G. and Tong, L. (2006), “Type based estimation over multi access channels”, *IEEE Trans. Sign. Proc.*, **54**(2), 613-626.

Moghim, N., Safavi, S.M. and Hashemi, M.R. (2010), “Performance evaluation of a new end-point admission control algorithm in NGN with improved network utilization”, *Int. J. Innov. Comput. Inf. Control*, **6**(7), 3067-3080.

Newman, M.E.J. and Watts D.J. (1999), “Renormalization group analysis of the small-world network model”, *Phys. Lett.* A, **263**(4), 341-346.

Ribeiro, A. and Giannakis, G.B. (2006), “Bandwidth-constrained distributed estimation for wireless sensor networks, Part I: Gaussian case”, *IEEE Trans. Sign. Proc.*, **54**(3), 1131-1143.

Sayed, A.H. and Lopes, C.G. (2006), “Distributed recursive least-squares strategies over adaptive networks”, *Proceeding of Asilomar Conference Signals, Systems, Computers*, 233-237.

Strogatz S.H. (2001), “Exploring complex networks”, *Nature*, **410**(6825), 268-276.

Sung, W. and Chen, C. (2010), “Parallel data fusion for an industrial automatic monitoring system using radial basis function networks”, *Int. J. Innov. Comput. Inf. Control*, **1**(6), 2523-2536.

Touri, B. and Nedic, A. (2009), “Distributed consensus over network with noisy links”, *Proceeding of the 12th International Conference Information Fusion*, Seattle, WA, 146-154.

Tu, S.Y. and Sayed, A.H. (2011), “Mobile adaptive networks”, *IEEE J. Sel. Topics. Sign. Proc.*, **5**(4), 649-664.

Wang, X.F. (2002), “Complex networks: topology, dynamics and synchronization”, *Int. J. Bifurcation Chaos*, **12**(5), 885-916.

Willsky, A.S., Bello, M., Castanon, D.A., Levy, B.C. and

Verghese, G. (1982), "Combining and updating of local estimates and regional maps along sets of one-dimensional tracks", *IEEE Trans. Automa. Control*, **27**(4), 799-813.
Ying Liu, Chunguang Li, Wallace K.S. Tang and Zhaoyang Z. (2012), "Distributed estimation over complex networks", *Inform. Sci.*, **197**, 91-104.

CY