

A Variable Step-Size Diffusion Normalized Least-Mean-Square Algorithm with a Combination Method Based on Mean-Square Deviation

Sang Mok Jung · Ji-Hye Seo · PooGyeon Park

Received: 29 August 2014 / Revised: 10 February 2015 / Accepted: 11 February 2015 /
Published online: 25 February 2015
© Springer Science+Business Media New York 2015

Abstract A novel diffusion normalized least-mean-square algorithm is proposed for distributed network. For the adaptation step, the upper bound of the mean-square deviation (MSD) is derived instead of the exact MSD value, and then, the variable step size is obtained by minimizing it to achieve fast convergence rate and small steady-state error. For the diffusion step, the individual estimate at each node is constructed via the weighted sum of the intermediate estimates at its neighbor nodes, where the weights are designed by using a proposed combination method based on the MSD at each node. The proposed MSD-based combination method provides effective weights by using the MSD at each node as a reliability indicator. Simulations in a system identification context show that the proposed algorithm outperforms other algorithms in the literatures.

Keywords Adaptive filters · Distributed network · Distributed estimation · Diffusion normalized least-mean-square algorithm · Variable step size · Combination method

S. M. Jung (✉) · P. Park
Department of Electrical Engineering, Pohang University of Science and Technology,
Gyungbuk 790-784, Korea
e-mail: illus2@postech.ac.kr

P. Park
e-mail: ppg@postech.ac.kr

J.-H. Seo
Department of Division of IT Convergence Engineering, Pohang University of Science
and Technology, Gyungbuk 790-784, Korea
e-mail: bmclubhs@postech.ac.kr

1 Introduction

Adaptive filtering algorithms are widely used for signal processing in practical systems [1, 8, 9, 16]. Recently, diffusion adaptive filtering algorithms over distributed network have been widely studied [4, 5, 12, 19, 20]. These algorithms for distributed network are useful in many applications such as wireless sensor networks, distributed cooperative sensing and so on [6, 11].

In the adaptive filtering algorithms, the step size governs the convergence rate and the steady-state error. To meet the conflicting requirements of fast convergence rate and small steady-state error, the step size needs to be controlled. In the literatures, various schemes for controlling the step size have been proposed [2, 17], and they are expanded to diffusion adaptive filtering algorithm recently [7, 10, 14, 15].

In the diffusion adaptive filtering algorithms, a set of nodes in the network cooperates to estimate an unknown system. That is, each node of the network can share and fuse information at its neighbors through a combination method. The combination method determines weights connecting each node to its neighbor nodes. Therefore, the performance of the diffusion algorithms depends a lot on the design of the combination method. From this point of view, many algorithms mainly focused on designing the combination method by utilizing the network topology and/or statistics such as the signal-to-noise ratio (SNR) conditions. General combination methods are shown in Table 1. Among them, the uniform [4] and the metropolis [20] methods are based solely on topology of the distributed network. Therefore, they are sensitive to the network characteristics such as SNR conditions. To determine the weights by considering the network characteristics, the relative degree-variance method [5] is proposed. In this method, the estimate of every neighbor is weighted proportionally according to its inverse of noise variance. Thus, combination weights for the nodes that has the larger measurement noise will be smaller than other nodes.

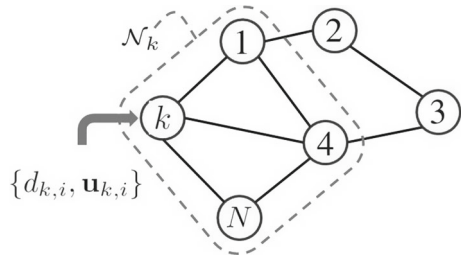
This paper proposes a variable step size scheme and a combination method by analyzing the mean-square deviation (MSD) for the diffusion normalized least-mean-square (D-NLMS) algorithm. The MSD analysis is carried out based on [3, 13, 18]. For the adaptation step, the upper bound of the MSD is derived by analyzing the behavior of the MSD propagation instead of the exact MSD value because it cannot be obtained directly. The variable step size is obtained by minimizing the upper bound of the MSD at next iteration. The step size could be optimally controlled from the variable step size, and it yields improved performance in terms of the convergence

Table 1 Combination methods

Method	Combination coefficients $a_{l,k}$
Uniform [4]	$1/n_k$
Metropolis [20]	$1/\max(n_k, n_l)$
Relative degree-variance [5]	$\frac{n_l \sigma_{v,l}^{-2}}{\sum_{m \in \mathcal{N}_k} n_m \sigma_{v,m}^{-2}}$

In all cases, $a_{l,k} = 0$ if $l \notin \mathcal{N}_k$, and $a_{k,k}$ is chosen such that $\sum_{l=1}^N a_{l,k} = 1$ for all k

Fig. 1 A distributed network with N nodes. At every i , node k takes a measurement $\{d_{k,i}, \mathbf{u}_{k,i}\}$ and generates an estimate $\mathbf{w}_{k,i}$ in cooperation with its neighbor set \mathcal{N}_k



rate and the steady-state error. The upper bound of the MSD and the optimal step size are reciprocally updated from each others at every iteration. For the diffusion step, a new combination method based on the MSD is proposed to overcome drawbacks of existing combination methods. The MSD at each node is used as its own reliability indicator because it represents preciseness of estimation and contains information that how much the estimate at each node is reliable. In the proposed MSD-based combination method, the combination weights can be determined adaptively because the MSD at each node is iteratively updated by using the recursion of the upper bound of the MSD that is derived in the adaptation step. Therefore, it provides more intuitive and effective weights than existing methods. The proposed variable step size and the MSD-based combination method provide performance improvement of the algorithm. The performance of the proposed algorithm is verified by simulations.

2 Diffusion Least-Mean-Square Algorithms

Consider a distributed network with N nodes (Fig. 1). The global system has an FIR vector $\mathbf{w}^o \in \mathcal{R}^{M \times 1}$. At every node k and time instant i , the desired signal $d_{k,i}$ is represented as

$$d_{k,i} = \mathbf{u}_{k,i}^T \mathbf{w}^o + v_{k,i}, \tag{1}$$

where $\mathbf{u}_{k,i}$ denotes the white Gaussian input vector; $v_{k,i}$ is the measurement noise that is assumed to be identically distributed and statistically independent of the input vector. It is assumed that $v_{k,i}$ has a zero-mean white Gaussian distribution with variance $\sigma_{v,k}^2$.

The adapt-then-combine (ATC) scheme [5] has been introduced in the literature for the diffusion least-mean-square (D-LMS) algorithm. By assuming that only intermediate estimates at each node are exchanged without their measurement data, the D-LMS algorithm is given as following two steps : the adaptation and the diffusion steps.

In the adaptation step, each node updates an intermediate estimate as

$$\boldsymbol{\psi}_{k,i} = \mathbf{w}_{k,i-1} + \mu_{k,i} \mathbf{u}_{k,i} \left(d_{k,i} - \mathbf{u}_{k,i}^T \mathbf{w}_{k,i-1} \right), \tag{2}$$

where $\mathbf{w}_{k,i}$ is an individual estimate of \mathbf{w}^o , $\boldsymbol{\psi}_{k,i}$ is an intermediate estimate, and $\mu_{k,i}$ is a local step size at node k . In the diffusion step, each node computes an individual estimate from weighted sum of intermediate estimates of its neighbor nodes as

$$\mathbf{w}_{k,i} = \sum_{l=1}^N a_{l,k} \boldsymbol{\psi}_{l,i}. \quad (3)$$

The $a_{l,k}$ is the weight connecting node k to its neighbor node $l \in \mathcal{N}_k$, where \mathcal{N}_k is set of neighbor nodes for node k including itself and n_k is the degree at node k (the number of neighbor nodes for node k). The weight $a_{l,k}$ can be formed using several combination methods in Table 1.

3 Variable Step Size Diffusion Normalized Least-Mean-Square Algorithm with MSD-Based Combination Method

The D-NLMS algorithm can be easily formulated as an extension of the D-LMS algorithm which solves its sensitivity to the input data. Based on the conventional normalized least-mean-square (NLMS) algorithm [8] with consideration about the distributed network, the D-NLMS algorithm is given by

(Adaptation Step)

$$\boldsymbol{\psi}_{k,i} = \mathbf{w}_{k,i-1} + \mu_{k,i} \frac{\mathbf{u}_{k,i} (d_{k,i} - \mathbf{u}_{k,i}^T \mathbf{w}_{k,i-1})}{\mathbf{u}_{k,i}^T \mathbf{u}_{k,i}}, \quad (4)$$

(Diffusion Step)

$$\mathbf{w}_{k,i} = \sum_{l=1}^N a_{l,k} \boldsymbol{\psi}_{l,i}. \quad (5)$$

In this section, the variable step size is derived for the adaptation step to establish the fast convergence rate and small steady-state estimation errors. Furthermore, for the diffusion step, a new combination method interprets the MSD as a reliability indicator and adopts it to determine the weights for the intermediate estimates.

3.1 The Adaptation Step : MSD Analysis and Variable Step Size

In the adaptation step, the step size $\mu_{k,i}$ is adjusted by minimizing the MSD which is defined as

$$\hat{p}_{k,i} \triangleq E \left(\tilde{\boldsymbol{\psi}}_{k,i}^T \tilde{\boldsymbol{\psi}}_{k,i} | \mathcal{U}_{k,i} \right) = \text{Tr}(\hat{\mathbf{P}}_{k,i}), \quad (6)$$

$$p_{k,i} \triangleq E \left(\tilde{\mathbf{w}}_{k,i}^T \tilde{\mathbf{w}}_{k,i} | \mathcal{U}_{k,i} \right) = \text{Tr}(\mathbf{P}_{k,i}), \quad (7)$$

where $\tilde{\boldsymbol{\psi}}_{k,i} = \mathbf{w}^o - \boldsymbol{\psi}_{k,i}$ and $\tilde{\mathbf{w}}_{k,i} = \mathbf{w}^o - \mathbf{w}_{k,i}$ are the intermediate and the individual weight error vectors, $\mathbf{P}_{k,i} = E \left(\tilde{\mathbf{w}}_{k,i} \tilde{\mathbf{w}}_{k,i}^T | \mathcal{U}_{k,i} \right)$, $\hat{\mathbf{P}}_{k,i} = E \left(\tilde{\boldsymbol{\psi}}_{k,i} \tilde{\boldsymbol{\psi}}_{k,i}^T | \mathcal{U}_{k,i} \right)$, $\mathcal{U}_{k,i} = \{\mathbf{u}_{k,j} | 0 \leq j < i\}$, and $\text{Tr}(\cdot)$ is the trace of the matrix. The update Eq. (4) can be described in terms of $\tilde{\boldsymbol{\psi}}_{k,i}$ and $\tilde{\mathbf{w}}_{k,i}$ as

$$\tilde{\boldsymbol{\psi}}_{k,i} = \mathbf{F}_{k,i} \tilde{\mathbf{w}}_{k,i-1} - \mu_{k,i} v_{k,i} \mathbf{u}_{k,i} \left(\mathbf{u}_{k,i}^T \mathbf{u}_{k,i} \right)^{-1}, \tag{8}$$

where

$$\mathbf{F}_{k,i} = \left(\mathbf{I} - \mu_{k,i} \mathbf{u}_{k,i} \left(\mathbf{u}_{k,i}^T \mathbf{u}_{k,i} \right)^{-1} \mathbf{u}_{k,i}^T \right). \tag{9}$$

To describe the transient behavior of the MSD, the input vector $\mathbf{u}_{k,i}$ is considered as a given quantity because $\mathbf{P}_{k,i}$ is a conditional expectation associated with $\mathbf{u}_{k,i}$. If the dependencies of $\tilde{\boldsymbol{\psi}}_{k,i-1}$ and $v_{k,i}$ are neglected, $\hat{\mathbf{P}}_{k,i}$ is obtained as follows:

$$\hat{\mathbf{P}}_{k,i} = \mathbf{F}_{k,i} \mathbf{P}_{k,i-1} \mathbf{F}_{k,i}^T + \mu_{k,i}^2 \sigma_{v,k}^2 \mathbf{u}_{k,i} \left(\mathbf{u}_{k,i}^T \mathbf{u}_{k,i} \right)^{-2} \mathbf{u}_{k,i}^T. \tag{10}$$

Then, recursion of the MSD is derived by taking the trace of both sides of Eq. (10) as follows:

$$\text{Tr} \left(\hat{\mathbf{P}}_{k,i} \right) = \text{Tr} \left(\mathbf{F}_{k,i}^T \mathbf{F}_{k,i} \mathbf{P}_{k,i-1} \right) + \mu_{k,i}^2 \sigma_{v,k}^2 \left(\mathbf{u}_{k,i}^T \mathbf{u}_{k,i} \right)^{-1}. \tag{11}$$

In this recursion, $\text{Tr} \left(\mathbf{F}_{k,i}^T \mathbf{F}_{k,i} \mathbf{P}_{k,i-1} \right)$ satisfies the relation according [13] as

$$\begin{aligned} & (-2\mu_{k,i} + \mu_{k,i}^2) \lambda_1(\mathbf{P}_{k,i-1}) + \text{Tr}(\mathbf{P}_{k,i-1}), \\ & \leq \text{Tr} \left(\mathbf{F}_{k,i}^T \mathbf{F}_{k,i} \mathbf{P}_{k,i-1} \right) \\ & \leq (-2\mu_{k,i} + \mu_{k,i}^2) \lambda_M(\mathbf{P}_{k,i-1}) + \text{Tr}(\mathbf{P}_{k,i-1}), \end{aligned} \tag{12}$$

where $\lambda_j(\mathbf{P}_{k,i-1})$ is the j th largest eigenvalue of the $\mathbf{P}_{k,i-1}$. Since $\lambda_M(\mathbf{P}_{k,i-1}) \leq \text{Tr}(\mathbf{P}_{k,i-1})/M$, there exists a positive constant $\beta \geq 1$ such that

$$\lambda_M(\mathbf{P}_{k,i-1}) \simeq \frac{\text{Tr}(\mathbf{P}_{k,i-1})}{\beta M}. \tag{13}$$

For white Gaussian input case, all diagonal elements of $\mathbf{P}_{k,i-1}$ have the same value because the elements in the input vector are independent to each other. Therefore, $\mathbf{P}_{k,i-1}$ can be simplified with only one element $p_{k,i-1}/M$ such as $p_{k,i-1}/M \times \mathbf{I}$. Then, above equation can be simply rewritten as

$$\lambda_M(\mathbf{P}_{k,i-1}) = \frac{p_{k,i-1}}{M}. \tag{14}$$

That is, the value of β can be set to 1 for the white Gaussian input case. On the other hand, when the input vector is correlated, the condition number of matrix $\mathbf{P}_{k,i-1}$ will be worse. So, differences have arisen between the eigenvalues. In this case, β should be properly selected to compensate the effects of the bad condition number of $\mathbf{P}_{k,i-1}$. From above discussion and inequality in (12), Eq. (11) can be written as

$$\hat{p}_{k,i} \leq \left(1 - \frac{2\mu_{k,i} - \mu_{k,i}^2}{\beta M} \right) p_{k,i-1} + \frac{\mu_{k,i}^2 \sigma_{v,k}^2}{\mathbf{u}_{k,i}^T \mathbf{u}_{k,i}}. \tag{15}$$

From Eq. (15), the upper bound of the intermediate MSD is recursively determined by using the individual MSD at previous iteration. By minimizing the upper bound of $\widehat{p}_{k,i}$ in (15) with respect to $\mu_{k,i}$, the variable step size that most largely decreases the MSD of the intermediate estimate can be chosen. According to

$$\frac{\partial \widehat{p}_{k,i}}{\partial \mu_{k,i}} = \frac{2\mu_{k,i} - 2}{\beta M} p_{k,i-1} + \frac{2\mu_{k,i} \sigma_{v,k}^2}{\mathbf{u}_{k,i}^T \mathbf{u}_{k,i}}, \quad (16)$$

the variable step size for the best performance at node k and time instant i is obtained as follows:

$$\mu_{k,i} = \frac{p_{k,i-1} \mathbf{u}_{k,i}^T \mathbf{u}_{k,i}}{p_{k,i-1} \mathbf{u}_{k,i}^T \mathbf{u}_{k,i} + \beta M \sigma_{v,k}^2}. \quad (17)$$

The proposed step size contains regularization effects because of $\beta M \sigma_{v,k}^2$ term in the denominator. Therefore, the proposed algorithm can maintain its performance even for badly excited input signals. Furthermore, the step size can reduce the computational complexity because it is obtained by minimizing the scalar recursive Eq. (15) for the MSD instead of the matrix recursion. By using the proposed variable step size, the proposed algorithm can achieve fast convergence and small steady-state estimation error.

3.2 The Diffusion Step : MSD-Based Combination Method

In the diffusion step, the individual estimate at the node k is determined by fusing the intermediate estimates from neighbors of node k with combination methods. As mentioned in Sect. 1, however, general combination methods do not reflect the network characteristics or are too complicated to implement in real-time process. In this section, therefore, a new combination method based on the MSD at each node is proposed to improve preciseness of the individual estimate.

The intermediate MSD denotes how close the intermediate estimates are to the unknown system, so the inverse of $\widehat{p}_{l,i}$ can be used as the weight for $\psi_{l,i}$. To satisfy $\sum_{l=1}^N a_{l,k} = 1$ as a convex parameter, the weights are defined as

$$a_{l,k} \triangleq \begin{cases} \left(\sum_{m \in \mathcal{N}_k} \widehat{p}_{m,i}^{-1} \right)^{-1} \widehat{p}_{l,i}^{-1}, & \text{if } l \in \mathcal{N}_k \\ 0, & \text{otherwise} \end{cases} \quad (18)$$

In this proposed MSD-based combination method, the node k uses $\widehat{p}_{l,i}^{-1}$ as the reliability indicator for the node l . The key advantage of working with the inverse of MSD instead of conventional combination methods is that it directly provides information that node has good characteristics and how much the intermediate estimate at each node is reliable. This strategy allows each node to cooperate with only reliable information. For example, if node l has large MSD because of the large noise of the node l , then $\widehat{p}_{l,i}^{-1}$ becomes small. Thus, the weight for $\psi_{l,i}$ will be smaller than those of the other neighbor nodes. Consequently, the MSD-based combination method can effectively

determine which node is reliable to estimate the unknown system and will show more improved performance than existing combination methods. By fusing $\psi_{l,i}$ for $l \in \mathcal{N}_k$ through the MSD-based combination method in (18), $\mathbf{w}_{k,i}$ is determined as follows:

$$\begin{aligned} \mathbf{w}_{k,i} &= \sum_{l=1}^N a_{l,k} \psi_{l,i} \\ &= \sum_{l \in \mathcal{N}_k} \left(\sum_{m \in \mathcal{N}_k} \widehat{p}_{m,i}^{-1} \right)^{-1} \widehat{p}_{l,i}^{-1} \psi_{l,i}. \end{aligned} \tag{19}$$

To extract the relationship between the individual and the intermediate MSDs at current iteration, (5) is developed in terms of $\widetilde{\mathbf{w}}_{k,i}$ and $\widetilde{\psi}_{k,i}$ by subtracting \mathbf{w}^o from both sides as

$$\widetilde{\mathbf{w}}_{k,i} = \sum_{l \in \mathcal{N}_k} a_{l,k} \widetilde{\psi}_{l,i}. \tag{20}$$

Multiplying \mathbf{w}^{oT} to both sides of (20) and taking expectation lead to the following equation:

$$E \left(\widetilde{\mathbf{w}}_{k,i} \mathbf{w}^{oT} \right) = E \left(\sum_{l \in \mathcal{N}_k} a_{l,k} \widetilde{\psi}_{l,i} \mathbf{w}^{oT} \right). \tag{21}$$

Assume that the intermediate and the individual weight error vectors, $\widetilde{\psi}_{k,i}$ and $\widetilde{\mathbf{w}}_{k,i}$, are orthogonal to $\psi_{k,i}$ and $\mathbf{w}_{k,i}$, respectively, then the following relation is established:

$$-E \left(\widetilde{\mathbf{w}}_{k,i} \widetilde{\mathbf{w}}_{k,i}^T \right) = -E \left(\sum_{l \in \mathcal{N}_k} a_{l,k} \widetilde{\psi}_{l,i} \widetilde{\psi}_{l,i}^T \right), \tag{22}$$

which is represented as

$$\mathbf{P}_{k,i} = \sum_{l \in \mathcal{N}_k} a_{l,k} \widehat{\mathbf{P}}_{l,i}. \tag{23}$$

Taking trace to both sides of (23) leads to

$$\begin{aligned} p_{k,i} &= \sum_{l \in \mathcal{N}_k} a_{l,k} \widehat{p}_{l,i}, \\ &= \sum_{l \in \mathcal{N}_k} \left(\sum_{m \in \mathcal{N}_k} \widehat{p}_{m,i}^{-1} \right)^{-1} \widehat{p}_{l,i}^{-1} \widehat{p}_{l,i} \\ &= \frac{n_k}{\sum_{m \in \mathcal{N}_k} \widehat{p}_{m,i}^{-1}}, \end{aligned} \tag{24}$$

which is described with the harmonic sum of the intermediate MSD values. According to (24), $p_{k,i}$ is recursively updated by using $\widehat{p}_{l,i}$ where $l \in \mathcal{N}_k$. Subsequently, the

Table 2 The proposed algorithm

Initialization: $\psi_{k,0} = 0, \beta = 1, p_{k,0} = 10$ for all nodes k

for each i **do**

for each k **do**

Adaptation Step

$$\mu_{k,i} = \frac{p_{k,i-1} \mathbf{u}_{k,i}^T \mathbf{u}_{k,i}}{p_{k,i-1} \mathbf{u}_{k,i}^T \mathbf{u}_{k,i} + \beta M \sigma_{v,k}^2}$$

$$\hat{p}_{k,i} = \left(1 - \frac{2\mu_{k,i} - \mu_{k,i}^2}{\beta M} \right) p_{k,i-1} + \frac{\mu_{k,i}^2 \sigma_{v,k}^2}{\mathbf{u}_{k,i}^T \mathbf{u}_{k,i}}$$

$$\psi_{k,i} = \mathbf{w}_{k,i-1} + \frac{\mu_{k,i} \mathbf{u}_{k,i} (d_{k,i} - \mathbf{u}_{k,i}^T \mathbf{w}_{k,i-1})}{\mathbf{u}_{k,i}^T \mathbf{u}_{k,i}}$$

Diffusion Step

$$\mathbf{w}_{k,i} = \sum_{l \in \mathcal{N}_k} \left(\sum_{m \in \mathcal{N}_k} \hat{p}_{m,i}^{-1} \right)^{-1} \hat{p}_{l,i}^{-1} \psi_{l,i}$$

$$p_{k,i} = \frac{n_k}{\sum_{m \in \mathcal{N}_k} \hat{p}_{m,i}^{-1}}$$

end for

end for

reliability indicator at next iteration ($\hat{p}_{k,i+1}^{-1}$) is recursively updated according to (15) by using the $p_{k,i}$ at the present iteration which is updated in (24). The proposed algorithm is summarized in Table 2.

4 Simulation Results

The computer simulations are used to confirm that the proposed algorithm provides more precise estimates of \mathbf{w}^o than other algorithms. An unknown system is randomly generated with 16 taps ($M = 16$). The SNR for the measurement at node k is defined as follows:

$$\text{SNR}_k \triangleq 10 \log_{10} \frac{E(y_{k,i}^2)}{E(v_{k,i}^2)}, \quad (25)$$

where $y_{k,i} = \mathbf{u}_{k,i}^T \mathbf{w}^o$. The noise variance at the node k , $\sigma_{v,k}^2$, is considered as known value because it can be easily estimated during silences in practice. The MSD and the normalized mean squared deviation (NMSD) at node k can be computed as follows:

$$\text{MSD}_k \triangleq E \left(\|\mathbf{w}^o - \mathbf{w}_{k,i}\|^2 \right), \quad (26)$$

$$\text{NMSD}_k \triangleq E \left(\|\mathbf{w}^o - \mathbf{w}_{k,i}\|^2 / \|\mathbf{w}^o\|^2 \right). \quad (27)$$

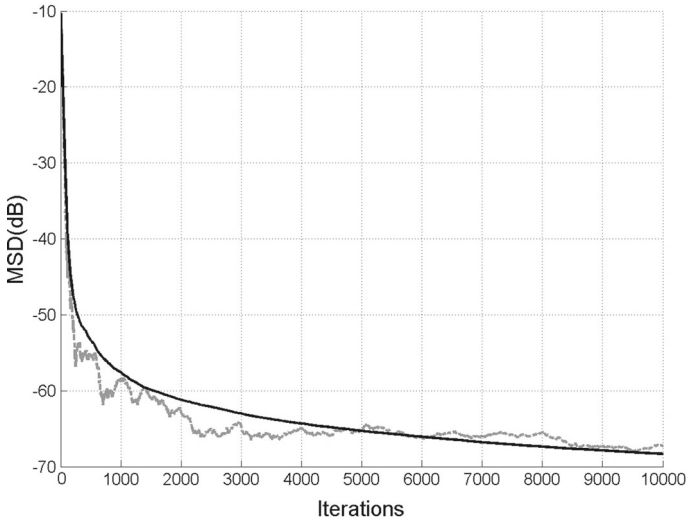


Fig. 2 Simulated MSD learning curve (*dashed gray line*) and its prediction (*solid black line*) based on the proposed analysis for distributed network with one node (SNR = 30 dB)

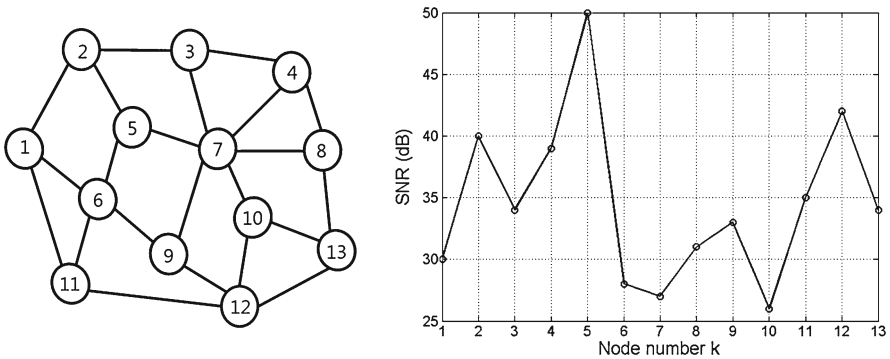


Fig. 3 A general distributed network with 13 nodes and their SNR values

The network NMSD is chosen as a performance indicator that is defined as follows:

$$\text{Network NMSD} \triangleq \frac{1}{N} \sum_{k=1}^N \text{NMSD}_k. \tag{28}$$

All simulation results are obtained by ensemble averaging over 100 trials. For fair comparison, the best tuning parameter values of the existing algorithms are determined after many trials. The relative degree-variance method is used in the diffusion step for the competing algorithms because it yields the best performance. Three distributed network topologies with the SNR values at each node are considered for simulations in Figs. 3, 5, and 7.

A Monte Carlo simulation is carried out in Fig. 2 to discuss regarding the accuracy of the proposed theoretical analysis to predict the MSD. Figure 2 demonstrates how

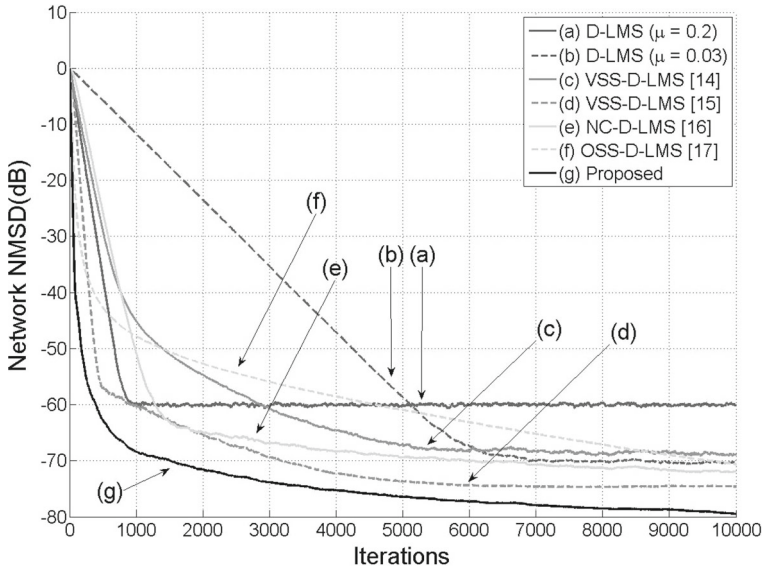


Fig. 4 NMSD learning curves of seven diffusion algorithms for the general distributed network in Fig. 3. $\alpha = 0.9$ [10], $\delta = 10^{-4}$ [10], $\alpha = 0.999$ [14], $\gamma = 0.01$ [14], $\alpha = 5 \times 10^{-4}$ [15], $\gamma = 300$ [15], $\epsilon = 10^{-6}$ [7], $\lambda = 0.8$ [7], and $\beta = 1$ for the proposed algorithm

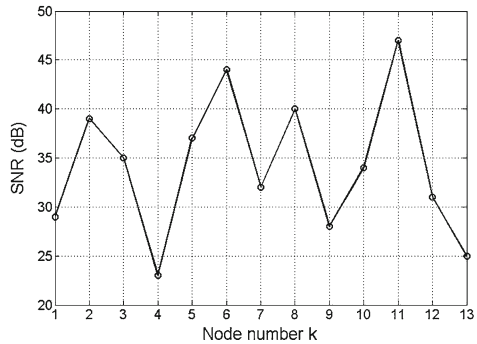
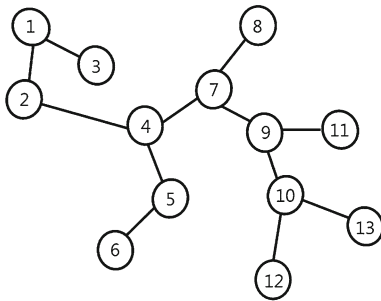


Fig. 5 A tree structure distributed network with 13 nodes and their SNR values. In this topology, there exist no cycle path

the proposed analysis predicts the MSD learning behavior of the D-NLMS. It shows that the proposed approach excellently predicts the learning behavior of the practice MSD. Therefore, proposed theoretical analysis of the MSD has enough accuracy to derive the optimal step size and the MSD-based combination method.

Figures 4 and 6 represent the network NMSD learning curves of the conventional D-LMS [5], the VSS-D-LMS [10], the VSS-D-LMS [14], the NC-D-LMS [15], the OSS-D-LMS [7], and the proposed algorithms with respect to the topologies in Figs. 3 and 5, respectively. The proposed algorithm achieves the smallest steady-state error and the fastest convergence rate among the competing algorithms, which means that

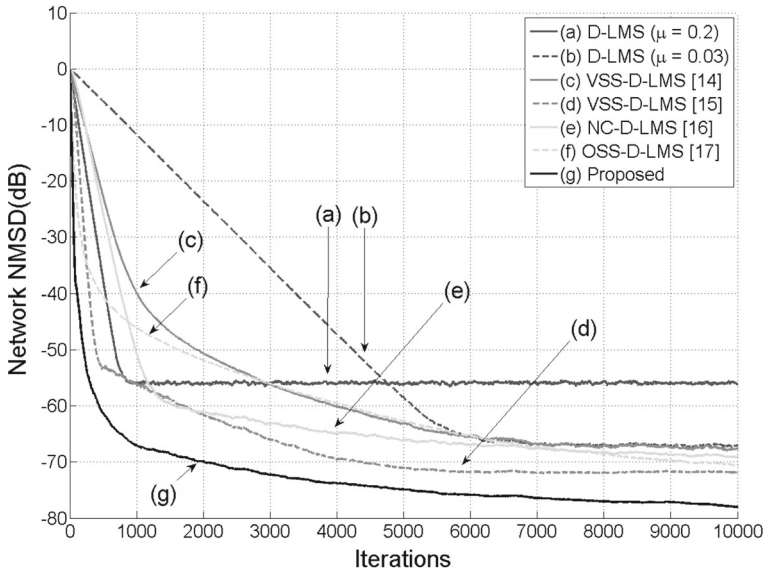


Fig. 6 NMSD learning curves of seven diffusion algorithms for the tree structure distributed network in Fig. 5. $\alpha = 0.95$ [10], $\delta = 10^{-5}$ [10], $\alpha = 0.999$ [14], $\gamma = 0.01$ [14], $\alpha = 5 \times 10^{-4}$ [15], $\gamma = 300$ [15], $\epsilon = 10^{-5}$ [7], $\lambda = 0.7$ [7], and $\beta = 1$ for the proposed algorithm

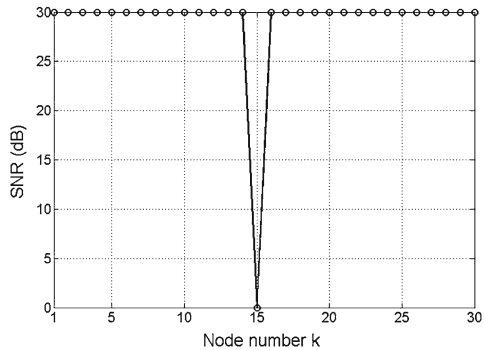
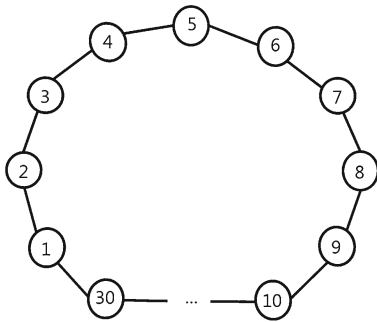


Fig. 7 An extremely bad distributed network with 30 nodes. At every node, k takes small measurement except node 15. As an interference, 0 dB measurement noise is added at the node 15

the step size derived in this paper is very close to the optimal value in the concern of the fast convergence rate on every iteration.

Figure 8 shows the network NMSD learning curves of the competing algorithms and the proposed algorithm with respect to the topology in Fig. 7. The network topology is ring structure, and all nodes are contaminated by white Gaussian noise with SNR=30dB except node 15. The node 15 contains extremely bad information with SNR=0dB, which may affect other nodes in the diffusion step. As can be seen, the proposed algorithm establishes the fastest convergence rate and the smallest steady-

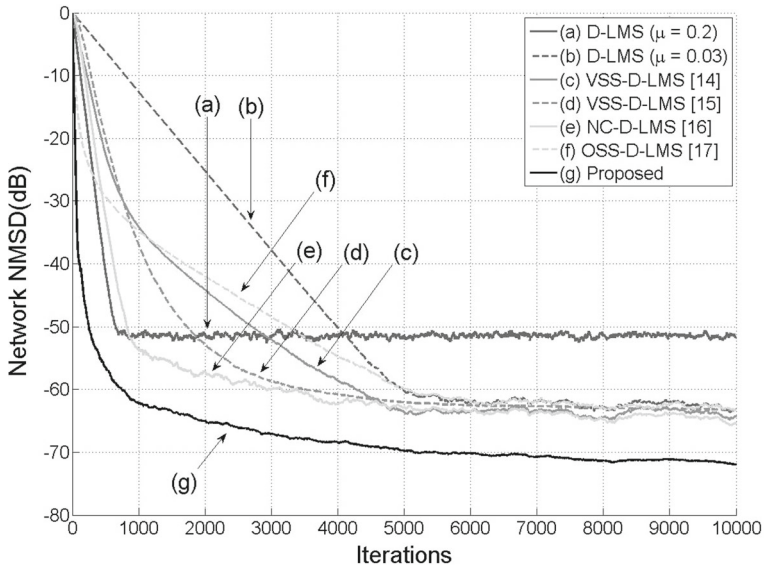


Fig. 8 NMSD learning curves of seven diffusion algorithms for the extremely bad distributed network in Fig. 7. $\alpha = 0.95$ [10], $\delta = 10^{-3}$ [10], $\alpha = 0.999$ [14], $\gamma = 10^{-3}$ [14], $\alpha = 7 \times 10^{-4}$ [15], $\gamma = 300$ [15], $\epsilon = 10^{-6}$ [7], $\lambda = 0.55$ [7], and $\beta = 1$ for the proposed algorithm

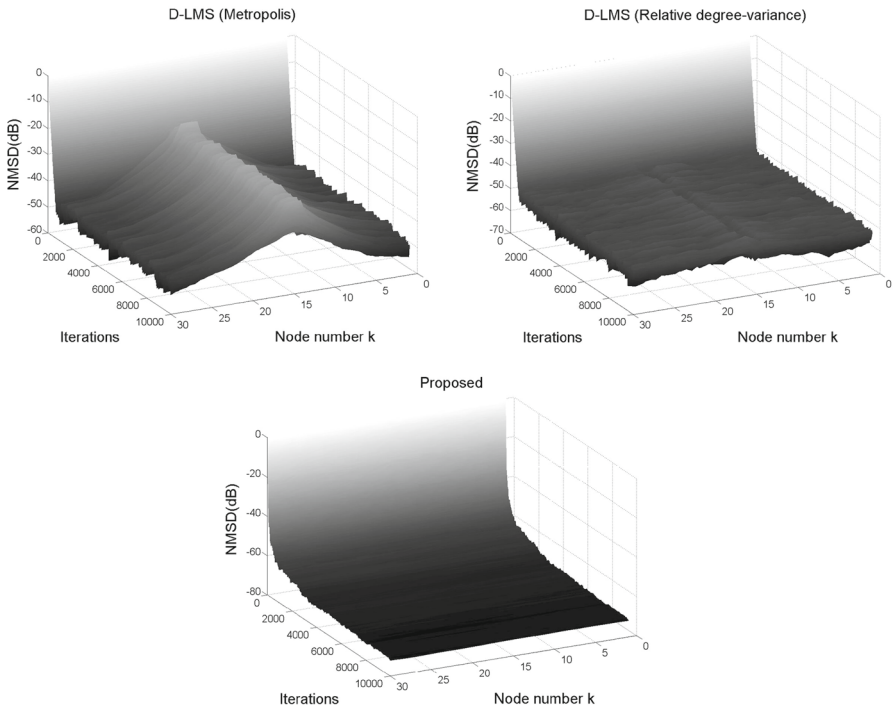


Fig. 9 NMSD learning curves of the D-LMS algorithm with different combination methods for the extremely distributed network in Fig. 7. $\mu = 0.2$ for the D-LMS, $\beta = 1$ for the proposed algorithm

state error, where the performances of the other algorithms are degraded due to bad information at node 15.

To confirm the effect of the proposed MSD-based combination method more clearly, an additional simulation result is represented in Fig. 9. It represents the NMSD learning curves of the conventional D-LMS algorithm with two different combination methods (metropolis and relative degree-variance) and the proposed algorithms. The D-LMS algorithm with the metropolis method is directly affected by bad information at node 15 because the combination is based only on the network topology. The D-LMS algorithm with the relative degree-variance method shows rather flat-shaped NMSD learning curves because it reflects the noise variance at each node as well as network topology when determining the combination coefficients to assign the smaller weight to the lower-SNR node. The NMSD learning curves of the proposed algorithm are clearly flat in spite of the bad information, and it can be said that the proposed variable step size scheme and the MSD-based combination method effectively suppress the effect of bad information.

5 Conclusion

In this paper, the variable step size and the MSD-based combination method were proposed for distributed networks. The variable step size derived by minimizing the upper bound of the MSD provided improvement of the filter performance in the aspects of the convergence rate and the steady-state estimation error. Furthermore, the MSD-based combination method can provide effective weights between the individual estimate and intermediate estimates by using the inverse of the MSD as a reliability indicator. The D-NLMS algorithm with the MSD-based combination method improved the performance against various network characteristics. Simulations showed that the proposed algorithm outperforms the existing diffusion algorithms.

Acknowledgments This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (NRF-2014R1A1A2055122).

References

1. S. Attallah, The wavelet transform-domain lms adaptive filter with partial subband-coefficient updating. *IEEE Trans. Circuits Syst. II* **53**(1), 8–12 (2006)
2. J. Benesty, H. Rey, L. Vega, S. Tressens, A nonparametric VSS NLMS algorithm. *IEEE Signal Process. Lett.* **13**(10), 581–584 (2006)
3. N.J. Bershad, Analysis of the normalized LMS algorithm with Gaussian inputs. *IEEE Trans. Acoust. Speech Signal Process. ASSP* **34**(4), 793–806 (1986)
4. V.D. Blondel, J. M. Hendrickx, A. Olshevsky, J. N. Tsitsiklis, Convergence in multi agent coordination, consensus, and flocking, in *Proceedings of Joint 44th IEEE Conference Decision Control European Control Conference (CDC-ECC)*, 2996–3000 (2005)
5. F.S. Cattivelli, A.H. Sayed, Diffusion LMS strategies for distributed estimation. *IEEE Trans. Signal Process.* **58**(3), 1035–1048 (2010)
6. Y. Chen, J. L. X. Yu, D.J. Hill, Multi-agent systems with dynamical topologies: consensus and applications. *IEEE Trans. Circuits Syst. Mag.* **13**(3), 21–34 (2013)
7. S. Ghazanfari-Rad and F. Labeau, Optimal variable step-size diffusion LMS algorithms, in *Proceedings IEEE SSP '14*, 464–467 (2014)

8. S. Haykin, *Adaptive Filter Theory*, 4th edn. (Prentice-Hall, New Jersey, 2002)
9. S.M. Jung, P. Park, Normalised least-mean-square algorithm for adaptive filtering of impulsive measurement noises and noisy inputs. *Electron. Lett.* **49**(20), 1270–1272 (2013)
10. H. S. Lee, S. E. Kim, J. W. Lee, W. J. Song, A variable step-size diffusion LMS algorithm for distributed estimation over networks, in *International Technical Conference on Circuits/Systems, Computers and Communications (ITC-CSCC)*, (2012)
11. D. Li, K.D. Wong, Y.H. Hu, A.M. Sayed, Detection, classification, and tracking of targets. *IEEE Signal Process. Mag.* **19**(2), 17–29 (2002)
12. C.G. Lopes, A.H. Sayed, Diffusion least-mean-squares over adaptive networks: Formulation and performance analysis. *IEEE Trans. Signal Process.* **56**(7), 3122–3136 (2008)
13. P. Park, M. Chang, N. Kong, Scheduled-stepsize NLMS algorithm. *IEEE Signal Process. Lett.* **16**(12), 1055–1058 (2009)
14. M.O.B. Saeed, A. Zerguine, S.A. Zummo, A variable step-size strategy for distributed estimation over adaptive networks, *EURASIP J. Adv. Signal Process.* **1**(135), 1–14 (2013)
15. M.O.B. Saeed, A. Zerguine, S.A. Zummo, A noise-constrained algorithm for estimation over distributed networks. *Int. J. Adapt. Control Signal Process.* **27**(10), 827–845 (2013)
16. A.H. Sayed, *Fundamental of Adaptive Filtering* (Wiley, New Jersey, 2003)
17. H.-C. Shin, A. Sayed, W.-J. Song, Variable step-size NLMS and affine projection algorithms. *IEEE Signal Process. Lett.* **11**(2), 132–135 (2004)
18. D.T.M. Slock, On the convergence behavior of the LMS and the normalized LMS algorithms. *IEEE Trans. Signal Process.* **41**(9), 2811–2825 (1993)
19. N. Takahashi, I. Yamada, A.H. Sayed, Diffusion least-mean-squares with adaptive combiners: Formulation and performance analysis. *IEEE Trans. Signal Process.* **58**(9), 4795–4810 (2010)
20. L. Xiao, S. Boyd, Fast linear iterations for distributed averaging. *Syst. Control Lett.* **53**(1), 65–78 (2004)