Signal Processing 140 (2017) 77–86

Contents lists available at ScienceDirect

Signal Processing

journal homepage: www.elsevier.com/locate/sigpro

Diffusion leaky LMS algorithm: Analysis and implementation

Lu Lu, Haiquan Zhao*

School of Electrical Engineering, Southwest Jiaotong University, Chengdu, China

ARTICLE INFO

Article history:
Received 14 December 2016
Revised 17 April 2017
Accepted 12 May 2017
Available online 12 May 2017

Keywords:
Distributed adaptation
Leaky LMS
Variable leakage factor
Acoustic echo cancellation (AEC)

ABSTRACT

The diffusion least-mean-square (dLMS) algorithms have attracted much attention owing to its robustness for distributed estimation problems. However, the performance of such algorithms may change when they are implemented for acoustic echo cancellation (AEC) systems. To overcome this problem, a leaky dLMS algorithm is proposed in this work, which is characterized by its numerical stability and small steady-state error for noisy speech signals. Then, we perform some stability and convergence analyses of the proposed algorithm for Gaussian inputs and verify the theory results by simulations. As an added contribution in this paper, we further develop a new variable leakage factor (VLF) strategy for the leaky dLMS algorithm to overcome the parameter selection of adaptation. Finally, implementations of the proposed algorithms in the context of system identification and *stereophonic AEC (SAEC)* network are performed. Simulation results illustrate that the leaky diffusion algorithms achieve improved performance as compared with the existing algorithms.

1. Introduction

Owing to its simple structure and low computational cost, the least-mean-square (LMS) algorithm and its variant are widely used in adaptive signal processing [1–4]. However, in practice, it is well known that direct implementation of the conventional LMS algorithm can be problematic [5,6]. In such a case, the leaky LMS algorithm was proposed [6,7]. It can effectively solve the following problems: (a) Numerical problem. The problem is due to the inadequacy of excitation in the input data. Filter parameters obtain arbitrarily large values and may fail to work. (b) Stagnation behavior. This is due to the low input signal. Since the gradient estimate is too small to adjust the coefficients of the algorithm, the adaptation of the algorithm may stall [6,7]. The leaky-based algorithms have been applied in diverse fields [6–16], including active noise control (ANC) [12,13], channel estimation [15], and nonlinear acoustic echo cancellation (NLAEC) [16]. Moreover, some improved versions of the leaky LMS algorithm were proposed by introducing the leaky factor into cost function to improve stability or convergence rate [17–20].

Recently, to estimate some parameters of interest from the data collected at nodes distributed over a geographic region, several distributed estimation algorithms were developed, which have also been applied in many areas [21–27]. In the previous studies, the incremental strategy [28–31] and the diffusion strategy [32–37] of cooperation for adaptive networks have been studied extensively. In the incremental strategy, the definition of a cyclic path over the nodes is required, and this technique is sensitive to link failures [35]. The diffusion method, in contrast, is a more widely used strategy for distributed estimation. In this strategy, each node communicates with a subset of its neighbors, and can achieve robust performance and a stable behavior over networks regardless of the topology. Based on the variable step size scheme, some improvements have been reported to either enhance the convergence or to improve the accuracy of estimation [38–41]. In [42–48], some works on reduced-complexity diffusion-based adaptive estimation algorithms were proposed, which reduce the internode communications and retain the benefits of cooperation. Particularly, in [37], two versions of the diffusion LMS (dLMS) algorithm, the adapt-then-combine (ATC) scheme and the combine-then-adapt (CTA) scheme were proposed, based on the different orders of *adaptation* and *combination* steps. Note that the ATC version of dLMS algorithm outperforms the CTA version of dLMS algorithm in all cases, and better performance can be achieved if the measurements are shared [37]. Due to its merit, the ATC version of the dLMS algorithm (ATC dLMS) has been introduced to the subband adaptive filter [49], information theoretic learning (ITL) [50,51] to further obtain improved performance.

A distributed incremental leaky LMS has been developed in [52] that can be used to estimate the coefficients of network. Although this algorithm offers a stable performance, it cannot be applied to large networks which may prohibit its practical appli-

* Corresponding author.
E-mail addresses: lulu19900303@126.com, lulu@my.swjtu.edu.cn (L. Lu), hqzhao_swjtu@126.com (H. Zhao).

http://dx.doi.org/10.1016/j.sigpro.2017.05.015

Table 1
Mathematical notations.

Notations	Description
E(·)	Mathematical expectation
$(\cdot)^T$	Transposition
diag(·)	Diagonal matrix with entries given by the arguments
Tr{ · }	Trace operator
$\lambda_{\max}\{\cdot\}$	The largest eigenvalue of a matrix
$\|\|\cdot\|\|_p$	Matrix l_p norm of its argument
$\|\cdot\|$	Absolute value of a scalar
$\|\cdot\|^{-1}$	Inverse of a matrix
col{ · }	Column vector with scalar entries or vector entries
vec{ · }	Stack the columns of its matrix argument on top of each other
$\otimes$	Kronecker product
$\{\cdot\}^o$	Real value of parameter
I	Identity matrix
1	$N \times 1$ vector with unit entries

cations. More importantly, it needs to establish a *Hamiltonian cycle* over the network, which is an NP-hard problem, and it may not exist in the general case [53]. To overcome these limitations, a leaky dLMS algorithm with two combination strategies-ATC and CTA is proposed, which is motivated by recent work in [37,52]. Compared with the state-of-art algorithms, the leaky dLMS algorithm is derived by minimizing the instantaneous leaky objective function rather than the mean square error (MSE) cost function, and it has a superiority performance via the ATC strategy. In addition, we provide the stochastic behavior and stability analyses of the leaky dLMS algorithm for Gaussian inputs.

Many variable leakage factor (VLF) algorithms have been proposed in the literature [54–56], where the main objective is to improve the convergence speed of the algorithm. In [54], Subudhi *et al.* attempted to investigate whether VLF scheme can be adopted to develop an efficient strategy for power system frequency estimation. We observe an improved performance of this effort, and investigate employing such VLF scheme to the leaky dLMS algorithm for distributed estimation.

In summary, our main contributions are listed as follows: (1) to develop a novel leaky dLMS algorithm that is suited for AEC applications, (2) to analyze the performance of the leaky dLMS algorithm in terms of the network mean square deviation (MSD) and stability behavior, (3) to develop a variable leaky dLMS algorithm in which the leakage factor is self-adjusting. The variable leaky dLMS algorithm remains the convergence rate of the leaky dLMS algorithm and overcomes the parameter selection problem in practice, and (4) we compare the simulation results of the proposed methods with the existing algorithms in the context of stereophonic AEC (SAEC) systems.

The rest of the paper is organized as follows. In Section 2, we derive the proposed leaky dLMS algorithm. In Section 3, the convergence analysis of the proposed algorithm is performed. In Section 4, we propose the VLF-based leaky dLMS algorithm and analyze the computational complexity. We show the computer simulation results in Section 5 and conclude the paper in Section 6.

The notations used in this paper are listed in Table 1. We use normal letters to denote scalars, use boldface lowercase letters to denote vectors, and use boldface uppercase letters to denote the matrix.

2. Proposed algorithm

Consider a network of N sensor nodes distributed over a geographic area. At each time instant i, each sensor node $k \in \{1, 2, \ldots, N\}$ has access to the realization of some zero-mean random process $\{d_k(i), \boldsymbol{u}_{k,i}\}$, where $d_k(i)$ is the desired signal, and $\boldsymbol{u}_{k,i} = [u_{k,i}, u_{k,i-1}, \ldots, u_{k,i-M+1}]^T$ is a $1 \times M$ regression vector. Suppose these measurements at every node follow a standard model given by:

$$d_k(i) = \boldsymbol{u}_{k,i}^T \boldsymbol{w}^o + v_k(i) \tag{1}$$

where $\boldsymbol{w}^o$ is the unknown parameter vector, and $v_k(i)$ is the measurement noise with variance $\sigma_{v,k}^2$. Here, we assume that $v_k(i)$ and $\boldsymbol{u}_{k,i}$ are spatially independent and independent identically distributed (i.i.d.), and $v_k(i)$ is independent of $\boldsymbol{u}_{k,i}$.

Define the error signal at node k as

$$e_{k,i} = d_k(i) - \boldsymbol{u}_{k,i}^T \boldsymbol{w}_{k,i-1} \tag{2}$$

where $\boldsymbol{w}_{k,i-1}$ is the estimate of $\boldsymbol{w}^o$ at node k and time $i-1$.

The update equation of ATC dLMS can be expressed as [37]

$$\begin{cases} \boldsymbol{\varphi}_{k,i} = \boldsymbol{w}_{k,i-1} + \mu \boldsymbol{u}_{k,i}^T\left(d_k(i) - \boldsymbol{u}_{k,i}^T \boldsymbol{w}_{k,i-1}\right) \\ \boldsymbol{w}_{k,i} = \sum_{l \in \mathcal{N}_k} a_{l,k} \boldsymbol{\varphi}_{l,i} \end{cases} \tag{3}$$

where μ is the step size (learning rate), $\boldsymbol{\varphi}_{k,i}$ is the local estimates at node k and time i, and $\mathcal{N}_k$ is the set of nodes with which node k shares information (including k itself). The weighting coefficients $\{a_{l,k}\}$ are the real, non-negative, satisfying the condition $a_{l,k} = 0$ if $l \notin \mathcal{N}_k$. The notation $\mathbf{A}$ is the matrix with individual entries $\{a_{l,k}\}$

Inspired by the derivation of the dLMS algorithm [37], a novel algorithm for distributed parameter estimation, named leaky dLMS algorithm is proposed. Many well-known diffusion algorithms were derived by using gradient decent, including the dLMS algorithm [37] and the diffusion minimum error entropy algorithm [50]. Here we derive the leaky dLMS algorithm also by using gradient descent. For each node k, we seek an estimate of $\boldsymbol{w}^o$ by minimizing the following cost function:

$$\begin{aligned} J_k^{loc}(\boldsymbol{w}) &= \sum_{l \in \mathcal{N}_k} a_{l,k} \mathrm{E}|e_{l,i}|^2 + \gamma \boldsymbol{w}^T \boldsymbol{w} \\ &= \sum_{l \in \mathcal{N}_k} a_{l,k} \mathrm{E}|d_l(i) - \boldsymbol{u}_{l,i}^T \boldsymbol{w}|^2 + \gamma \boldsymbol{w}^T \boldsymbol{w} \end{aligned} \tag{4}$$

where $\gamma > 0$ is the leaky factor, and $\boldsymbol{w}$ is the estimate of $\boldsymbol{w}^o$.

Using the steepest-descent method, we have

$$\nabla_{\boldsymbol{w}} J_k^{loc}(\boldsymbol{w}) = \sum_{l \in \mathcal{N}_k} a_{l,k} \frac{\partial \{\mathrm{E}|e_{l,i}|^2 + \gamma \boldsymbol{w}^T \boldsymbol{w}\}}{\partial \boldsymbol{w}}. \tag{5}$$

Therefore, the updating of proposed algorithm for estimating $\boldsymbol{w}^o$ at node k is given by

$$\boldsymbol{w}_{k,i} = \boldsymbol{w}_{k,i-1} - \mu \sum_{l \in \mathcal{N}_k} a_{l,k} \left.\frac{\partial \{\mathrm{E}|e_{l,i}|^2 + \gamma \boldsymbol{w}^T \boldsymbol{w}\}}{\partial \boldsymbol{w}}\right|_{\boldsymbol{w}_{k,i-1}}. \tag{6}$$

Under the linear combination assumption [35,37], the linear combination $\boldsymbol{w}_{k,i}$ at node k can be defined as

$$\boldsymbol{w}_{k,i-1} = \sum_{l \in \mathcal{N}_k} a_{l,k} \boldsymbol{\varphi}_{l,i-1}. \tag{7}$$

Introducing (7) to (6), we can obtain an iterative formula for the intermediate estimate

$$\begin{aligned} \boldsymbol{\varphi}_{l,i} &= \boldsymbol{\varphi}_{l,i-1} - \mu \left.\frac{\partial \mathrm{E}|e_{l,i}|^2 + \gamma \boldsymbol{w}^T \boldsymbol{w}}{\partial \boldsymbol{w}}\right|_{\boldsymbol{w}_{k,i-1}} \\ &\approx \boldsymbol{\varphi}_{l,i-1} - \mu \left.\frac{\partial \mathrm{E}|e_{l,i}|^2 + \gamma \boldsymbol{w}^T \boldsymbol{w}}{\partial \boldsymbol{w}}\right|_{\boldsymbol{w}_{l,i-1}} \\ &= (1 - \gamma\mu)\boldsymbol{\varphi}_{l,i-1} + \mu \boldsymbol{u}_{l,i}^T e_{l,i}. \end{aligned} \tag{8}$$

An important point needs to be highlighted. The equation in (8) is not realizable, due to the fact that linear combination estimation $\boldsymbol{w}_{k,i-1}$, which requires to gradient calculation, is not available for node l [50]. Let's consider an approximation method as in [50], in which the difference between $\boldsymbol{w}_{k,i-1}$ and $\boldsymbol{w}_{l,i-1}$ should not be too large, owing to the fact that $\boldsymbol{w}_{k,i-1}$ and $\boldsymbol{w}_{l,i-1}$ are both linear-combination estimate of $\boldsymbol{w}^o$ at instant $i-1$.

Thus, we obtain the leaky dLMS algorithm by transforming the steepest-descent type iteration of (6) into a two-step iteration

$$\begin{cases} \boldsymbol{\varphi}_{k,i} = \boldsymbol{w}_{k,i-1} - \mu \frac{\partial \mathrm{E}|e_{k,i}|^2 + \gamma \boldsymbol{w}^T \boldsymbol{w}}{\partial \boldsymbol{w}}\Big|_{\boldsymbol{w}_{k,i-1}} & (adaptation) \\ \boldsymbol{w}_{k,i} = \sum_{l \in \mathcal{N}_k} a_{l,k} \boldsymbol{\varphi}_{l,i} & (combination). \end{cases} \tag{9}$$

In *adaptation* step of (9), the local estimate $\boldsymbol{\varphi}_{k,i-1}$ is replaced by linear combination $\boldsymbol{w}_{k,i-1}$. Such substitution is reasonable, because the linear combination contains more data information from neighbor nodes than $\boldsymbol{\varphi}_{k,i-1}$ [37]. Then, we extend the leaky dLMS algorithm to its ATC and CTA forms.

ATC leaky dLMS algorithm (without measurements exchange):

$$\begin{cases} \boldsymbol{\varphi}_{k,i} = (1-\mu\gamma)\boldsymbol{w}_{k,i-1} + \mu \boldsymbol{u}_{k,i}^T \big(d_k(i) - \boldsymbol{u}_{k,i}^T \boldsymbol{w}_{k,i-1}\big) \\ \boldsymbol{w}_{k,i} = \sum_{l \in \mathcal{N}_k} a_{l,k} \boldsymbol{\varphi}_{l,i}\ . \end{cases} \tag{10}$$

CTA leaky dLMS algorithm (without measurements exchange):

$$\begin{cases} \boldsymbol{\varphi}_{k,i-1} = \sum_{l \in \mathcal{N}_k} a_{l,k} \boldsymbol{w}_{l,i-1} \\ \boldsymbol{w}_{k,i} = (1-\mu\gamma)\boldsymbol{\varphi}_{k,i-1} + \mu \boldsymbol{u}_{k,i}^T \big(d_k(i) - \boldsymbol{u}_{k,i}^T \boldsymbol{\varphi}_{k,i-1}\big). \end{cases} \tag{11}$$

From [37], we observed that the ATC-based diffusion algorithms outperform the CTA-based versions in all cases. For this reason, we will focus on analyzing the ATC leaky dLMS algorithm in next section.

3. Stochastic analysis of the leaky dLMS algorithm

3.1. Mean performance

In this section, we will conduct the stochastic behavior analysis of the leaky dLMS algorithm, and formulate its stability condition. To carry out the analysis, we shall assume that:

Assumption 1. The input signals $\{\boldsymbol{u}_{k,\ i}\}$ are the zero-mean Gaussian signals, and are spatially and temporally independent;

Assumption 2. The noise signal $v_k(i)$ at each node k is assumed to be a Gaussian noise with zero-mean, and is independent of any other signals.

Note that the above assumptions are often used by the practitioner of adaptive signal processing. Although there are not truly in practical applications, these assumptions can simplify the analysis of adaptive filter. Before that, some results in classical references indicate that theory analyses agree with the simulation results by using such assumptions [37,57].

Then, we proceed to the stochastic behavior analysis of leaky dLMS algorithm. The weight error vector and intermediate weight error vector at agent k and time i are respectively defined as follows:

$$\tilde{\boldsymbol{w}}_{k,i} \triangleq \boldsymbol{w}^o - \boldsymbol{w}_{k,i} \tag{12}$$

$$\tilde{\boldsymbol{\varphi}}_{k,i} \triangleq \boldsymbol{w}^o - \boldsymbol{\varphi}_{k,i}. \tag{13}$$

Introduce the global quantities of the network weight vector $\boldsymbol{w}_i$, weight error vector $\tilde{\boldsymbol{w}}_i$ and intermediate network weight error vector $\tilde{\boldsymbol{\varphi}}_i$, i.e.,

$$\boldsymbol{w}_i \triangleq \mathrm{col}\{\boldsymbol{w}_{1,i}, \boldsymbol{w}_{2,i}, \ldots, \boldsymbol{w}_{N,i}\} \tag{14}$$

$$\tilde{\boldsymbol{w}}_i \triangleq \mathrm{col}\{\tilde{\boldsymbol{w}}_{1,i}, \tilde{\boldsymbol{w}}_{2,i}, \ldots, \tilde{\boldsymbol{w}}_{N,i}\} \tag{15}$$

$$\tilde{\boldsymbol{\varphi}}_i \triangleq \mathrm{col}\{\tilde{\boldsymbol{\varphi}}_{1,i}, \tilde{\boldsymbol{\varphi}}_{2,i}, \ldots, \tilde{\boldsymbol{\varphi}}_{N,i}\}. \tag{16}$$

The error vector is defined as

$$\boldsymbol{e}_i \triangleq \mathrm{col}\{e_{1,i}, e_{2,i}, \ldots, e_{N,i}\} \tag{17}$$

the noise vector is defined as

$$\boldsymbol{v}_i \triangleq \mathrm{col}\{v_1(i), v_2(i), \ldots, v_N(i)\} \tag{18}$$

the desired vector is defined as

$$\boldsymbol{d}_i \triangleq \mathrm{col}\{d_1(i), d_2(i), \ldots, d_N(i)\} \tag{19}$$

and introduce the diagonal matrix for step sizes at all nodes

$$\mathbf{M} \triangleq \mathrm{diag}\{\mu, \mu, \ldots, \mu\}. \tag{20}$$

Considering $\boldsymbol{e}_i \triangleq \boldsymbol{d}_i - \boldsymbol{w}_i^T \boldsymbol{u}_i$, the adaptation step in (10) can be written as

$$\begin{aligned} \tilde{\boldsymbol{\varphi}}_i &= \tilde{\boldsymbol{w}}_{i-1} - \mathbf{\Upsilon}\mathbf{Q}\tilde{\boldsymbol{w}}_{i-1}\boldsymbol{u}_i\boldsymbol{u}_i^T - \mathbf{Q}\mathbf{V}_i\boldsymbol{u}_i + \mathbf{Q}\mathbf{\Upsilon}\boldsymbol{w}_{i-1} \\ &= \big[\mathbf{I}_{MN} - \mathbf{Q}\big(\mathbf{\Upsilon} + \boldsymbol{u}_i\boldsymbol{u}_i^T\big)\big]\tilde{\boldsymbol{w}}_{i-1} - \mathbf{Q}\mathbf{V}_i\boldsymbol{u}_i + \mathbf{Q}\mathbf{\Upsilon}\boldsymbol{\mathcal{W}^o} \end{aligned} \tag{21}$$

where

$$\boldsymbol{\mathcal{W}^o} \triangleq \mathrm{col}\{\boldsymbol{w}^o, \boldsymbol{w}^o, \ldots, \boldsymbol{w}^o\} \tag{22}$$

$$\boldsymbol{u}_i \triangleq \mathrm{col}\{\boldsymbol{u}_{1,i}, \boldsymbol{u}_{2,i}, \ldots, \boldsymbol{u}_{N,i}\} \tag{23}$$

$$\mathbf{Q} = \mathbf{M} \otimes \mathbf{I}_M \tag{24}$$

$$\mathbf{\Upsilon} = \mathbf{\Upsilon}_s \otimes \mathbf{I}_M, \quad \mathbf{\Upsilon}_s \triangleq \mathrm{diag}\{\gamma, \gamma, \ldots, \gamma\} \tag{25}$$

and

$$\mathbf{V}_i = \mathbf{V}_s \otimes \mathbf{I}_M, \quad \mathbf{V}_s \triangleq \mathrm{diag}\{\boldsymbol{v}_i\}. \tag{26}$$

Combining with the combination step in (10), the update formulation of the weight deviation vector of the proposed algorithm can be expressed as:

$$\tilde{\boldsymbol{w}}_i = \mathbf{P}^T\big[\mathbf{I}_{MN} - \mathbf{Q}(\mathbf{\Upsilon} + \boldsymbol{u}_i\boldsymbol{u}_i^T)\big]\tilde{\boldsymbol{w}}_{i-1} - \mathbf{P}^T\mathbf{Q}\mathbf{V}_i\boldsymbol{u}_i + \mathbf{P}^T\mathbf{Q}\mathbf{\Upsilon}\boldsymbol{\mathcal{W}^o} \tag{27}$$

where

$$\mathbf{P} = \mathbf{A} \otimes \mathbf{I}_M. \tag{28}$$

Taking expectations of both sides of (27), we get

$$\begin{aligned} \mathrm{E}\{\tilde{\boldsymbol{w}}_i\} = {} & \mathbf{P}^T[\mathbf{I}_{MN} - \mathbf{Q}(\mathbf{\Upsilon} + \boldsymbol{u}_i\boldsymbol{u}_i^T)]\mathrm{E}\{\tilde{\boldsymbol{w}}_{i-1}\} \\ & -\mathbf{P}^T\mathbf{Q}\mathrm{E}\{\mathbf{V}_i\boldsymbol{u}_i\} + \mathbf{P}^T\mathbf{Q}\mathbf{\Upsilon}\boldsymbol{\mathcal{W}^o}. \end{aligned} \tag{29}$$

Under Assumption 1, (29) can be approximated as

$$\mathrm{E}\{\tilde{\boldsymbol{w}}_i\} = \mathbf{P}^T\big[\mathbf{I}_{MN} - \mathbf{Q}(\mathbf{\Upsilon} + \boldsymbol{u}_i\boldsymbol{u}_i^T)\big]\mathrm{E}\{\tilde{\boldsymbol{w}}_{i-1}\} + \mathbf{P}^T\mathbf{Q}\mathbf{\Upsilon}\boldsymbol{\mathcal{W}^o}. \tag{30}$$

Therefore, the weight deviation vector $\mathrm{E}\{\tilde{\boldsymbol{w}}_i\}$ converges if $\lambda_{\max}\{\mathbf{P}^T[\mathbf{I}_{MN} - \mathbf{Q}(\mathbf{\Upsilon} + \boldsymbol{u}_i\boldsymbol{u}_i^T)]\} < 1$, where $\lambda_{\max}$ stands for the largest eigenvalue of a matrix in absolute form. According to (30), the general form is asymptotically unbiased for any initial condition and any choice of matrix $\mathbf{P}^T$. Hence, the problem reduces to determine the stability condition for the matrix $\mathbf{I}_{MN} - \mathbf{Q}(\mathbf{\Upsilon} + \boldsymbol{u}_i\boldsymbol{u}_i^T)$. We notice that the matrix $\mathbf{I}_{MN} - \mathbf{Q}(\mathbf{\Upsilon} + \boldsymbol{u}_i\boldsymbol{u}_i^T)$ is a block-diagonal matrix, so it can be easily verified that it is stable if its block-diagonal entries $\mathbf{I}_M - \mu(\gamma + \boldsymbol{u}_{k,i}\boldsymbol{u}_{k,i}^T)$ are stable. Consequently, the vector in leaky dLMS algorithm converges only if

$$0 < \mu < \frac{2}{\gamma + \lambda_{\max}\{\boldsymbol{u}_{k,i}\boldsymbol{u}_{k,i}^T\}} \tag{31}$$

Therefore in the steady state, i.e., $n \to \infty$, we have

$$\begin{aligned}\mathcal{W}^o - \mathrm{E}\{\boldsymbol{w}_\infty\} =& \mathbf{P}^T[\mathbf{I}_{MN} - \mathbf{Q}(\boldsymbol{\Upsilon} + \boldsymbol{\Lambda})]\big[\mathcal{W}^o - \mathrm{E}\{\boldsymbol{w}_\infty\}\big] - \mathbf{P}^T\mathbf{Q}\boldsymbol{\Upsilon}\mathcal{W}^o \\ =& \big(\mathbf{P}^T[\mathbf{I}_{MN} - \mathbf{Q}(\boldsymbol{\Upsilon} + \boldsymbol{\Lambda})] - \mathbf{P}^T\mathbf{Q}\boldsymbol{\Upsilon}\big)\mathcal{W}^o \\ & - \mathbf{P}^T[\mathbf{I}_{MN} - \mathbf{Q}(\boldsymbol{\Upsilon} + \boldsymbol{\Lambda})]\mathrm{E}\{\boldsymbol{w}_\infty\}\end{aligned} \tag{32}$$

where $\boldsymbol{\Lambda}$ is the matrix of the eigenvalues, $\boldsymbol{\Lambda} = \mathrm{diag}\{\lambda_1, \lambda_2, \ldots, \lambda_{MN}\}$. Then, (32) can be reduced to

$$\begin{aligned}\mathrm{E}\{\boldsymbol{w}_\infty\} =& \Big\{\mathbf{P}^T[\mathbf{I}_{MN} - \mathbf{Q}(\boldsymbol{\Upsilon} + \boldsymbol{\Lambda})] - \mathbf{I}_{MN}\Big\}^{-1} \\ & \times\Big\{\mathbf{P}^T[\mathbf{I}_{MN} - \mathbf{Q}(\boldsymbol{\Upsilon} + \boldsymbol{\Lambda})] - \mathbf{P}^T\mathbf{Q}\boldsymbol{\Upsilon} - \mathbf{I}_{MN}\Big\}\mathcal{W}^o.\end{aligned} \tag{33}$$

For $\boldsymbol{\Upsilon} = \mathbf{0}$ ($\gamma = 0$), the mean convergence condition for the ATC leaky LMS algorithm is the same as that for the LMS based solution. In other words, the mean convergence condition is independent of the leaky factor γ. Yet, the bias depends on γ, and a small bias is achieved when γ is small. Therefore, a small γ should be adopted for the simulations. The above analysis is a suitable measure for describing the algorithm convergence and often used to evaluate the algorithm performance. However, the convergence in mean does not guarantee convergence of MSE. Therefore, we need to perform the necessary and sufficient condition of leaky dLMS algorithm in mean square sense.

3.2. Mean-square performance

This section is devoted to analyzing mean-square performance of the leaky dLMS algorithm for a white Gaussian input. We also follow the assumption 1 and 2 that are suitable for this analysis. First, the network mean-square deviation (NMSD) is defined as the average of MSD value over the network, i.e.,

$$\xi(i) = \frac{1}{N}\sum_{k=1}^{N}\xi_k(i) \tag{34}$$

where ξ_k is the MSD at node k. Multiplying (27) by its transpose and then taking the expectation, we obtain

$$\begin{aligned}&\mathrm{E}\{\tilde{\boldsymbol{w}}_i\tilde{\boldsymbol{w}}_i^T\} = \\ &\mathbf{P}^T[(\mathbf{I}_{MN} - \mathbf{Q}\boldsymbol{\Upsilon} - \mathbf{Q}\boldsymbol{u}_i\boldsymbol{u}_i^T)\mathrm{E}\{\tilde{\boldsymbol{w}}_{i-1}\tilde{\boldsymbol{w}}_{i-1}^T\}(\mathbf{I}_{MN} - \mathbf{Q}\boldsymbol{\Upsilon} - \mathbf{Q}\boldsymbol{u}_i\boldsymbol{u}_i^T)]\mathbf{P} \\ &+\mathbf{P}^T\mathbf{Q}\mathbf{V}_i\boldsymbol{u}_i\boldsymbol{u}_i^T\mathbf{V}_i^T\mathbf{Q}^T\mathbf{P} + \mathbf{P}^T\mathbf{Q}\boldsymbol{\Upsilon}\mathcal{W}^o\mathcal{W}^{oT}\boldsymbol{\Upsilon}^T\mathbf{Q}^T\mathbf{P} \\ &+\mathbf{P}^T\mathbf{Q}\boldsymbol{\Upsilon}\mathcal{W}^o\mathrm{E}\{\tilde{\boldsymbol{w}}_{i-1}^T(\mathbf{I}_{MN} - \mathbf{Q}\boldsymbol{\Upsilon} - \mathbf{Q}\boldsymbol{u}_i\boldsymbol{u}_i^T)\}\mathbf{P} \\ &+\mathbf{P}^T\mathrm{E}\{(\mathbf{I}_{MN} - \mathbf{Q}\boldsymbol{\Upsilon} - \mathbf{Q}\boldsymbol{u}_i\boldsymbol{u}_i^T)\tilde{\boldsymbol{w}}_{i-1}\}\mathcal{W}^{oT}\boldsymbol{\Upsilon}^T\mathbf{Q}^T\mathbf{P}.\end{aligned} \tag{35}$$

Note that (35) is obtained by eliminating the following terms:

$$\mathrm{E}\Big\{\mathbf{P}^T\mathbf{Q}\mathbf{V}_i\boldsymbol{u}_i\mathcal{W}^{oT}\boldsymbol{\Upsilon}^T\mathbf{Q}^T\mathbf{P}\Big\} = 0 \tag{36}$$

$$\mathrm{E}\Big\{\mathbf{P}^T\mathbf{Q}\boldsymbol{\Upsilon}\mathcal{W}^o\boldsymbol{u}_i\mathbf{V}_i\mathbf{Q}^T\mathbf{P}\Big\} = 0 \tag{37}$$

$$\mathrm{E}\Big\{(\mathbf{I}_{MN} - \mathbf{Q}\boldsymbol{\Upsilon} - \mathbf{Q}\boldsymbol{u}_i\boldsymbol{u}_i^T)\tilde{\boldsymbol{w}}_{i-1}\boldsymbol{u}_i^T\mathbf{V}_i^T\mathbf{Q}^T\mathbf{P}\Big\} = 0 \tag{38}$$

$$\mathrm{E}\Big\{\mathbf{P}^T\mathbf{Q}\mathbf{V}_i\boldsymbol{u}_i\tilde{\boldsymbol{w}}_{i-1}^T(\mathbf{I}_{MN} - \mathbf{Q}\boldsymbol{\Upsilon} - \mathbf{Q}\boldsymbol{u}_i\boldsymbol{u}_i^T)\Big\} = 0. \tag{39}$$

These terms are zero because of Assumption 1 and 2. Taking the trace of both sides of (35), we find the relation given by

$$\begin{aligned}&\mathrm{tr}\Big(\mathrm{E}\{\tilde{\boldsymbol{w}}_i\tilde{\boldsymbol{w}}_i^T\}\Big) = \\ &\mathbf{P}^T[(\mathbf{I}_{MN} - \mathbf{Q}\boldsymbol{\Upsilon} - \mathbf{Q}\sigma_u^2)\mathrm{tr}\Big(\mathrm{E}\{\tilde{\boldsymbol{w}}_{i-1}\tilde{\boldsymbol{w}}_{i-1}^T\}\Big)(\mathbf{I}_{MN} - \mathbf{Q}\boldsymbol{\Upsilon} - \mathbf{Q}\sigma_u^2)]\mathbf{P} \\ &+\mathbf{P}^T\mathbf{Q}\{\sigma_u^2(MN\sigma_v^2)\}\mathbf{Q}^T\mathbf{P} + \mathbf{P}^T\mathbf{Q}\boldsymbol{\Upsilon}\mathcal{W}^o\mathcal{W}^{oT}\boldsymbol{\Upsilon}^T\mathbf{Q}^T\mathbf{P} \\ &+\mathbf{P}^T\mathbf{Q}\boldsymbol{\Upsilon}\mathcal{W}^o\mathrm{tr}\Big(\mathrm{E}\{\tilde{\boldsymbol{w}}_{i-1}^T(\mathbf{I}_{MN} - \mathbf{Q}\boldsymbol{\Upsilon} - \mathbf{Q}\sigma_u^2)\}\Big)\mathbf{P} \\ &+\mathbf{P}^T\mathrm{tr}\Big(\mathrm{E}\{(\mathbf{I}_{MN} - \mathbf{Q}\boldsymbol{\Upsilon} - \mathbf{Q}\sigma_u^2)\tilde{\boldsymbol{w}}_{i-1}\}\Big)\mathcal{W}^{oT}\boldsymbol{\Upsilon}^T\mathbf{Q}^T\mathbf{P}\end{aligned} \tag{40}$$

where σ_u^2 is the variance of the input signal, and σ_v^2 is the variance of the noise signal.

From (34), we have

$$\xi_k(i) = \mathrm{tr}\Big(\tilde{\mathbf{W}}_k(i)\Big) \tag{41}$$

where $\tilde{\mathbf{W}}_k(i)$ is the kth diagonal block-element of $\mathrm{E}\{\tilde{\boldsymbol{w}}_i\tilde{\boldsymbol{w}}_i^T\}$.

3.3. Stability of the algorithm

The matrix $\mathbf{P}$ has real, non-negative entries. For this reason, the stability of (27) is only related to $\tilde{\boldsymbol{w}}_i = [\mathbf{I}_{MN} - \mathbf{Q}(\boldsymbol{\Upsilon} + \boldsymbol{u}_i\boldsymbol{u}_i^T)]\tilde{\boldsymbol{w}}_{i-1} - \mathbf{Q}\mathbf{V}_i\boldsymbol{u}_i + \mathbf{Q}\boldsymbol{\Upsilon}\boldsymbol{w}^o$. In other words, to guarantee the stability in mean sense, we have

$$\tilde{\boldsymbol{w}}_{k,i} = [1 - \mu(\gamma + \boldsymbol{u}_{k,i}\boldsymbol{u}_{k,i}^T)]\tilde{\boldsymbol{w}}_{k,i-1} - \mu v_k(i)\boldsymbol{u}_{k,i} + \mu\gamma\boldsymbol{w}^o. \tag{42}$$

We define the state vector as [45]:

$$\boldsymbol{\Gamma}_{i-1} = \begin{bmatrix}\boldsymbol{\Gamma}_{a,i-1} \\ \boldsymbol{\Gamma}_{b,i-1}\end{bmatrix} \tag{43}$$

where $\boldsymbol{\Gamma}_{a,i-1} = \mathrm{tr}\{\mathrm{E}[\tilde{\boldsymbol{w}}_{i-1}\tilde{\boldsymbol{w}}_{i-1}^T]\}$, and $\boldsymbol{\Gamma}_{b,i-1} = \mathrm{E}[\tilde{\boldsymbol{w}}_{i-1}]$. Using (43), the state equation can be expressed as:

$$\boldsymbol{\Gamma}_i = \Im\boldsymbol{\Gamma}_{i-1} + \boldsymbol{\Pi} \tag{44}$$

where

$$\Im = \begin{bmatrix}\Im_a & \Im_b \\ \mathbf{0} & \Im_d\end{bmatrix} \tag{45}$$

and

$$\Im_a = 1 - 2\mu(\sigma_{k,u}^2 + \gamma) + \mu^2[(M - 1 + \eta_{k,u})\sigma_{k,u}^4 + 2\sigma_{k,u}^2\gamma + \gamma^2] \tag{46}$$

$$\Im_b = -2\mu\gamma\big[1 - \mu(\sigma_{k,u}^2 + \gamma)\big]\boldsymbol{w}^o \tag{47}$$

$$\Im_d = 1 - \mu(\sigma_{k,u}^2 + \gamma) \tag{48}$$

where

$$\sigma_{k,u}^2 = \mathrm{E}\Big\{u_{k,u}^2\Big\} \tag{49}$$

$$\eta_{k,u} = \frac{\mathrm{E}\{u_{k,u}^4\}}{\sigma_{k,u}^4} \tag{50}$$

$$\sigma_{k,v}^2 = \mathrm{E}\Big\{v_k^2(i)\Big\} \tag{51}$$

and

$$\boldsymbol{\Pi} = \begin{bmatrix}\mu^2\Big(M\sigma_{k,u}^2\Big)\sigma_{k,v}^2 + \gamma^2||\boldsymbol{w}^o||_2^2 \\ -\mu\gamma\boldsymbol{w}^o\end{bmatrix}. \tag{52}$$

For stability, a practical bound can be obtained from (44). Note that $\Im$ is a triangular matrix, it has M eigenvalues equal to $1 - \mu(\gamma + \sigma_{k,u}^2)$ and one eigenvalue equal to the first entry in the

Table 2
Computational complexity for node k per iteration.

Algorithms	Multiplications	Additions	Memory words
ATC dLMS	$2M+MN_k+2$	$2M+(M-1)N_k-1$	$(N_k+2)M+5$
ATC RZA dLMS	$5M+MN_k+2$	$4M+(M-1)N_k-1$	$(N_k+3)M+6$
VSS dLMS	$2M+MN_k+4$	$2M+(M-1)N_k$	$(N_k+2)M+7$
SV dLMS	$2M+MN_k+5$	$2M+(M-1)N_k$	$(N_k+2)M+7$
ATC leaky dLMS	$2M+MN_k+3$	$2M+(M-1)N_k$	$(N_k+3)M+7$
ATC variable leaky dLMS	$3M+2MN_k+5$	$2M+(M-1)N_k+1$	$(2N_k+2)M+7$

matrix. Obviously, the leaky dLMS algorithm can converge in the mean square if

$$\left|1-2\mu(\gamma+\sigma_{k,u}^2)+\mu^2[(M-1+\eta_{k,u})\sigma_{k,u}^4+2\sigma_{k,u}^2\gamma+\gamma^2]\right|<1. \tag{53}$$

Therefore, we can see that the leaky dLMS algorithm converges if, and only if

$$0<\mu<\frac{2(\gamma+\sigma_{k,u}^2)}{(M-1+\eta_{k,u})\sigma_{k,u}^4+2\sigma_{k,u}^2\gamma+\gamma^2}. \tag{54}$$

Eq. (54) provides a mean-square stability condition on the learning rate μ that ensures the convergence of the leaky dLMS algorithm. If we set the leakage coefficient $\gamma=0$, then we obtain a bound on μ to guarantee convergence of the dLMS algorithm. Moreover, (54) shows the negative correlation between bound of the step size and the filter length.

4. Variable leaky dLMS algorithm

The bottleneck of the leaky dLMS algorithm is it requires a suitable leaky factor selection. In practice, such parameter setting is hard to implement, and therefore these fixed-leaky factor algorithms may not work very reliably. On the other hand, it is reasonable for each node to be calculated by an independent leaky factor. If the same leaky factors are used, every weight vector of the leaky dLMS 'leaky' out when the input at one node is turned off (other nodes have persistent excitations)[1]. Since different nodes share the different input signals, we need to assign an individual-variable-leaky-factor for each node.

Inspired by the method in [54,55], the leaky factor is adapted based on the stochastic gradient rule. Replacing γ by variable regularization parameters $\gamma_{k,i}$ for $k=1,2,\dots,N$, we have

$$\gamma_{k,i+1}=\gamma_{k,i}-\frac{\theta}{2}\frac{\partial e_{k,i}^2}{\partial\gamma_{k,i-1}} \tag{55}$$

where θ is a positive parameter. The derivative term of (55) can be computed according to the chain rule:

$$\frac{\partial e_{k,i}^2}{\gamma_{k,i-1}}=\left[\frac{\partial e_{k,i}^2}{\partial\boldsymbol{w}_{k,i}}\right]^T\frac{\partial\boldsymbol{w}_{k,i}}{\partial\gamma_{k,i-1}}. \tag{56}$$

From (2) we have

$$\frac{\partial e_{k,i}^2}{\partial\boldsymbol{w}_{k,i}}=-2e_{k,i}\boldsymbol{u}_{k,i} \tag{57}$$

[1] When the input is turned off, the weight vector of the basic dLMS algorithm stalls. For this reason, the leaky algorithm reduces the effects of non-persistent excitation as compared with the dLMS algorithm.

and the second derivative in (56) can be obtained from (9) as follows:

$$\begin{aligned}\frac{\partial\boldsymbol{w}_{k,i}}{\partial\gamma_{k,i-1}}&=\frac{\partial\left\{\sum_{l\in\mathcal{N}_k}a_{l,k}\boldsymbol{\varphi}_{l,i}\right\}}{\partial\gamma_{k,i-1}}\\&=\frac{\partial\left\{\sum_{l\in\mathcal{N}_k}a_{l,k}\big[(1-\mu\gamma_{l,i-1})\boldsymbol{w}_{l,i-1}+\mu e_{l,i}\boldsymbol{u}_{l,i}\big]\right\}}{\partial\gamma_{k,i-1}}\\&=-\mu\sum_{l\in\mathcal{N}_k}a_{l,k}\boldsymbol{w}_{l,i-1}.\end{aligned} \tag{58}$$

Thus, using (58) and (57) in (56), and substituting (56) into (55), one can obtain

$$\gamma_{k,i+1}=\gamma_{k,i}-\mu\theta e_{k,i}\boldsymbol{u}_{k,i}\sum_{l\in\mathcal{N}_k}a_{l,k}\boldsymbol{w}_{l,k}. \tag{59}$$

Note that the adaptation (59) is derived based on the ATC leaky dLMS algorithm. Similarly, the variable leaky factor version of CTA leaky dLMS algorithm can be easily derived according to above-mentioned derivation. To avoid confusion, (59) is called the *ATC variable leaky dLMS* algorithm, and its CTA version is called *CTA variable leaky dLMS* algorithm.

Table 2 summarizes the computational complexity and memory requirement of the ATC leaky dLMS, ATC variable leaky dLMS, ATC reweighted zero-attracting (RZA) dLMS [58], variable step-size diffusion LMS (VSS dLMS) [38], sparse variable step size dLMS (SV dLMS) [40] and the ATC dLMS algorithms, where N_k denotes the number of components of the neighborhood set $\mathcal{N}_k$. As we can see, the ATC dLMS algorithm has the lowest computational complexity than other algorithms, and it demands $(N_k+2)M+5$ words. The variable step size version, VSS dLMS and SV dLMS algorithms, require more multiplications, additions and memory than the dLMS algorithm. The increase in computational cost of the VSS dLMS and SV dLMS algorithms compared with the dLMS algorithm is tolerable and can be compensated by the improved performance. The ATC RZA dLMS algorithm, slightly increases the computational complexity as compared with the ATC dLMS algorithm, it requires additional $M+1$ multiplications for computing zero attracting. Owing to using VLF scheme, the proposed ATC variable leaky dLMS algorithm has the highest computational complexity and memory requirement among these algorithms. It needs $3M+2MN_k+5$ multiplications, $2M+(M-1)N_k+1$ additions, and $(2N_k+2)M+7$ words for VLF scheme, still with an affordable computation burden.

The CTA-based algorithms have the same processing and communication complexity as the ATC-based algorithms, so we do not repeat them here.

5. Simulation results

In this section, we conduct a series of simulations to evaluate the performance of the proposed algorithms, including simulations on a system identification example and simulations on a SAEC network example. We compare the estimation results of the proposed

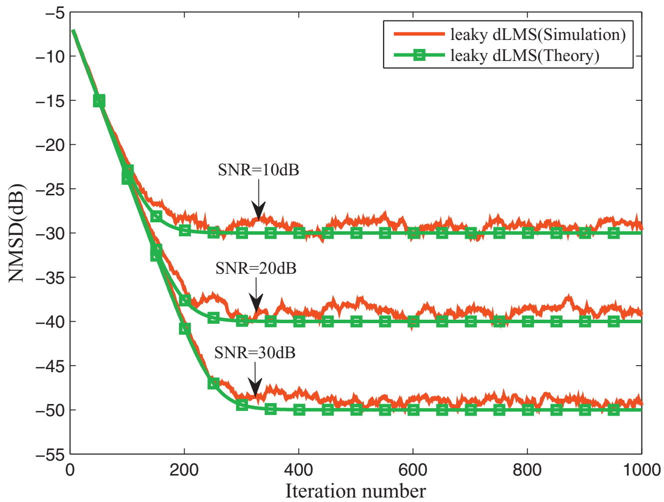

Fig. 1. A comparison of the theoretical NMSD in (40) with simulation results ($\mu = 0.02$, $\gamma = 0.0001$).

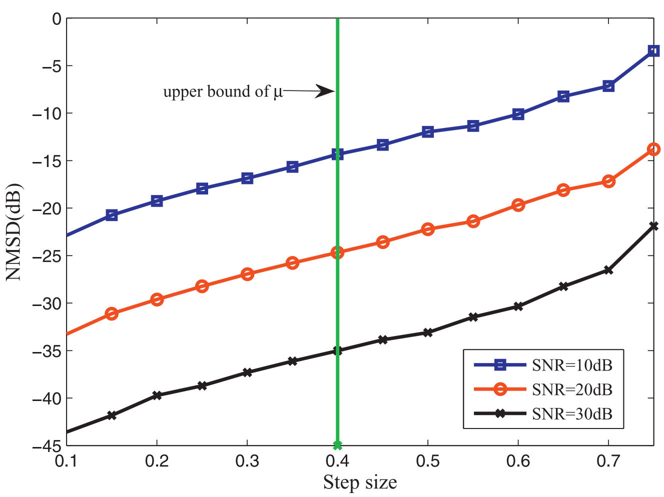

Fig. 2. NMSD of the proposed algorithm versus different step sizes.

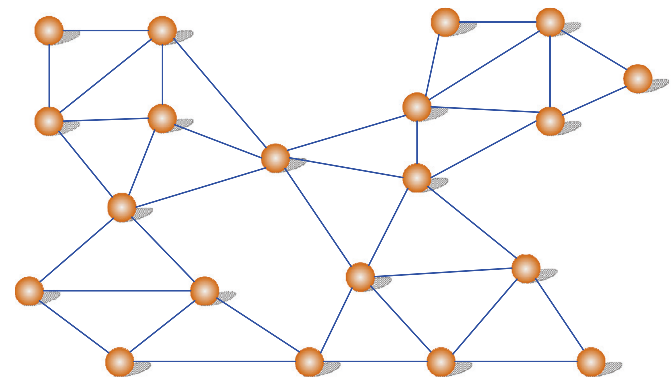

Fig. 3. Network topology.

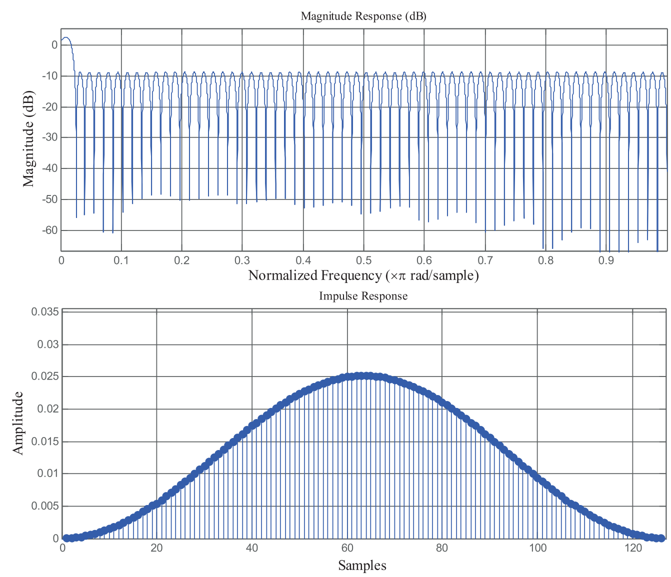

Fig. 4. Frequency response and impulse response of acoustic paths used in computer simulations.

algorithm with those of the dLMS algorithm, the VSS dLMS algorithm, the SV dLMS algorithm and the RZA dLMS algorithm. The performance of the algorithms is measured in terms of the NMSD, $\text{NMSD} = 10\log_{10}[\xi(i)]$.

Example 1: System identification with Gaussian input

In this subsection, we provide simulation verification for the analyses. The proposed algorithm for noise reduction network is composed of 4 nodes, and each node connects to its nearest 3 neighbor nodes. The unknown network vector has 5 entries. The Gaussian signal with zero mean and unit variance is employed as the input signal, $a_{l,k} = 1/n_k$ is used for $\{a_{l,\,k}\}$ [58,59], where n_k is the degree of node k.

Fig. 1 illustrates the NMSD curve of theoretical and simulation results for different signal-to-noise ratio (SNR) environments. In the leaky dLMS algorithm, the adaptation step size is $\mu = 0.02$, and the leaky factor is $\gamma = 0.0001$. As can be seen, the theoretical NMSDs agree with the simulation results in all cases. To further demonstrate the upper bound of the step size, Fig. 2 plots the steady-state NMSD of the leaky dLMS with different step sizes. It is found that the steady-state NMSD of the proposed algorithms increase as the step sizes increase. Besides, by using (54), we can obtain the theoretical upper bounds of step sizes for different nodes. The calculating data indicates that the upper bound of step size is about 0.4. To guarantee all the nodes work well, μ must locate in the range of $\mu < 0.4$ to ensure the stability.

Example 2: System identification with colored input

In second example, $a_{l,k} = 1/n_k$ was used for $\{a_{l,\,k}\}$ (uniform rule, [58]). The network is composed of 20 nodes, as shown in Fig. 3. The colored input signal $\boldsymbol{u}_{k,\,i}$ was generated by passing a zero-mean, white Gaussian noise (WGN) through a first-order system $T(z) = 1/(1-0.7z^{-1})$. The unknown vector of interest was modeled by a FIR filter of order $M = 128$[2], whose frequency response and impulse response were shown in Fig. 4. The variance of WGN and SNR over each node is illustrated in Fig. 5. All NMSD curves were obtained by ensemble averaging over 25 independent trials.

First, we investigate the performance of the algorithm using different γ values, as shown in Fig. 6. It can be easily observed that the CTA/ATC leaky dLMS algorithm is not very sensitive to this choice, but it turns out that the best option for ATC leaky dLMS algorithm is $\gamma = 0.01$. Correspondingly, the proper selection for CTA leaky dLMS algorithm is $\gamma = 0.001$. Fig. 7 displays a comparison of NMSD from the proposed algorithms and existing algorithms. One can see that the leaky dLMS algorithm outperforms the conventional dLMS algorithms in terms of steady-state NMSD. After the about iteration 1000, all the algorithms reach steady

[2] We also perform simulations for the proposed algorithms with $M = 5$, see http://arxiv.org/abs/1602.04329.

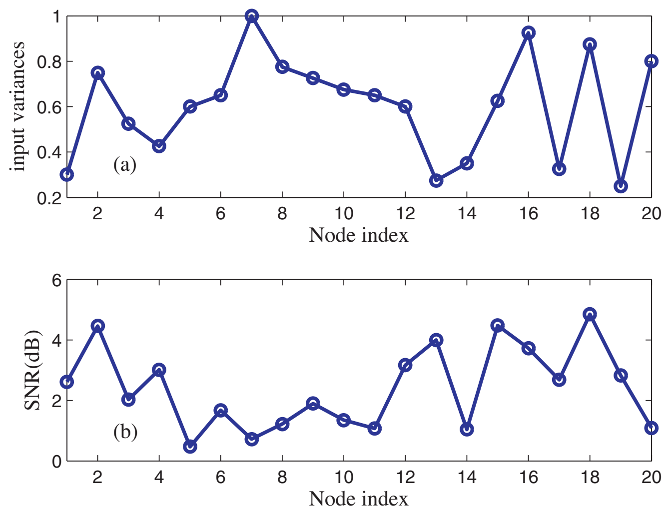

Fig. 5. Input variances and the SNR used in simulations.

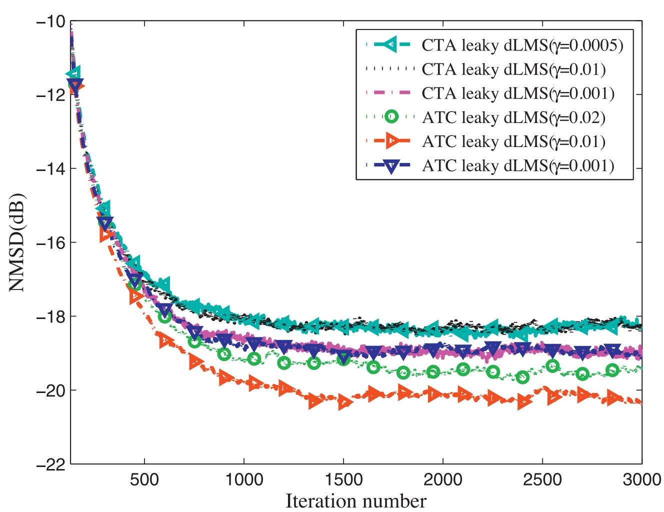

Fig. 6. NMSD of the proposed algorithms versus different γ ($\mu = 0.005$).

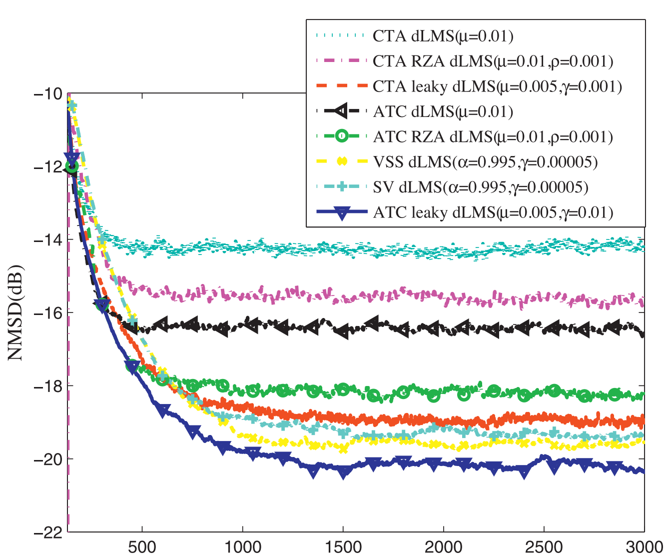

Fig. 7. NMSD of the proposed algorithms and the existing algorithms at SNR= 0 – 5dB.

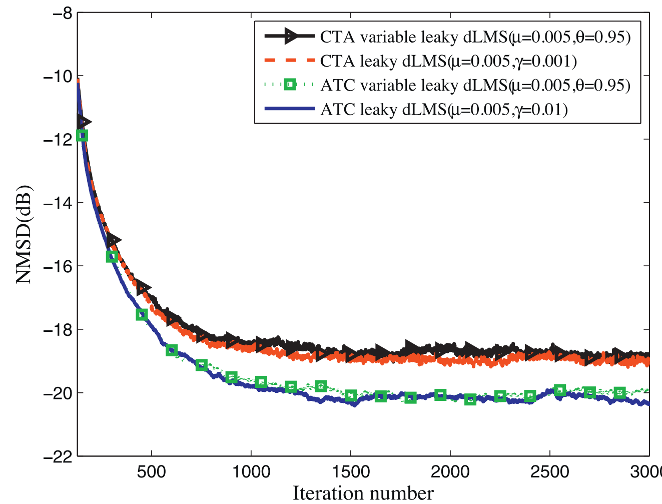

Fig. 8. NMSD of the proposed algorithms and there variable leaky versions at SNR= 0 – 5dB.

Table 3
Computation time.

Algorithms	Computation time(s)
ATC dLMS	1.29
ATC RZA dLMS	5.53
VSS dLMS	1.34
SV dLMS	2.87
CTA leaky dLMS	**8.32**
ATC leaky dLMS	**3.94**
ATC variable leaky dLMS	**5.69**

state. Besides, we found that the performance of two variable step-size diffusion algorithms is limited by the choice of their parameters α and γ. To obtain performance improvement and stability, we select α=0.95 and γ=0.00005 for the algorithms. In addition, it is obviously shown from these curves that the ATC leaky dLMS algorithm achieves low misadjustment (−20dB), whereas the other ATC-based algorithms finally settle at a NMSD value of about −14dB to −19dB. All the ATC-based algorithms achieve improved performance as compared with the CTA-based algorithms, with the similar initial convergence rate.

Next, we show the performance of variable leaky dLMS algorithms with the same simulation condition as Fig. 7. In Fig. 8 it can be seen that the performance of the variable leaky dLMS algorithms and the leaky dLMS algorithms are similar.

To evaluate the computational burden, we measured the average run execution time of the algorithm on a 2.1-GHz AMD processor with 8GB of RAM, running Matlab R2013a on Windows 7. The computation time of the algorithms is outlined in Table. 3. We see that the ATC-dLMS algorithm is the fastest one among these distributed algorithms, and the variable leaky dLMS algorithm slightly increases the execution time. The CTA algorithms achieve slower than the ATC algorithms, still with an affordable computation time.

Example 3: SAEC network

In a SAEC network simulation, we used four impulse responses of length $M = 128$ with sparseness, as shown in Fig. 9. The length of the adaptive filter was $2M = 256$. Note that the derivation of the proposed algorithm in SAEC network is different from Section 2, and therefore we show the derivation in Appendix A. The input signals were obtained by finite impulse response (FIR) filtering a common speech signal in the far-end location, and processing with positive and negative half-wave rectifiers (with $b = 0.5$) [60]. The output of the system is contaminated with 0dB SNR of WGN. To quantify the network performance of the algorithm, the network

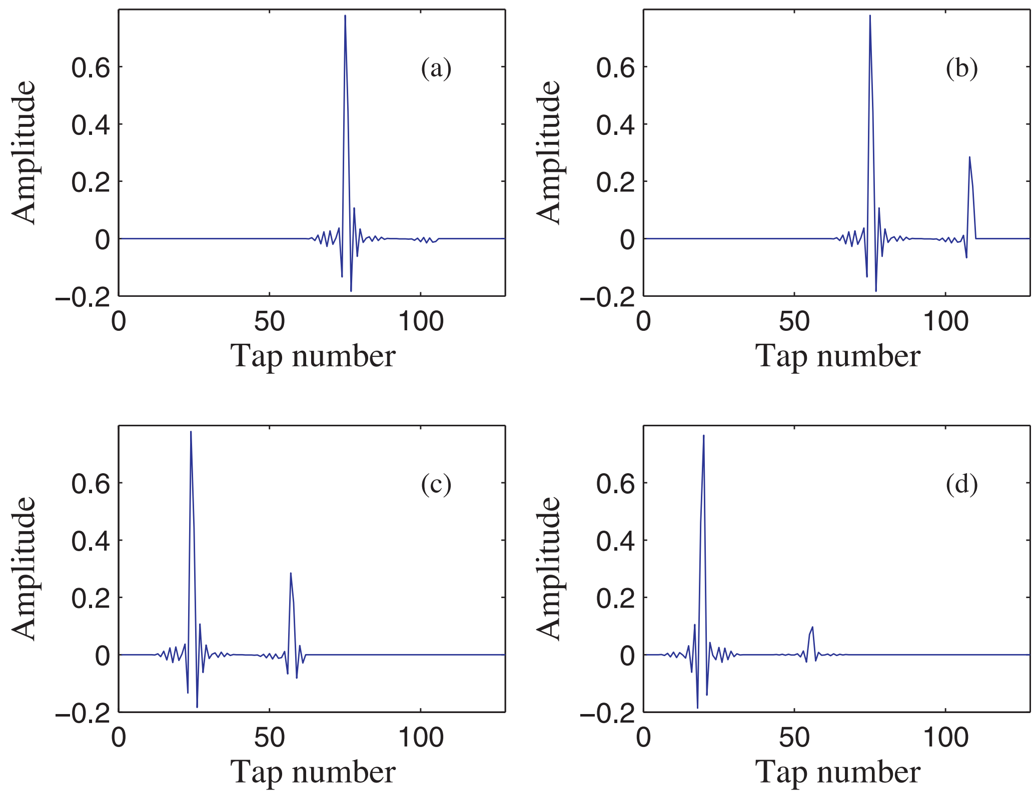

Fig. 9. Impulse responses used in SAEC network at node k. (a)$\boldsymbol{w}_{k,\,11}$ with sparsity 0.8853; (b)$\boldsymbol{w}_{k,\,12}$ with sparsity 0.8453; (c)$\boldsymbol{w}_{k,\,21}$ with sparsity 0.8327; (d)$\boldsymbol{w}_{k,\,22}$ with sparsity 0.8651, $k = 1, 2, , N$. The sparseness level of the responses is calculated according to $\zeta_s = \frac{M}{M-\sqrt{M}}\left(1 - \frac{||\boldsymbol{w}^o||_1}{\sqrt{M}||\boldsymbol{w}^o||_2}\right)$.

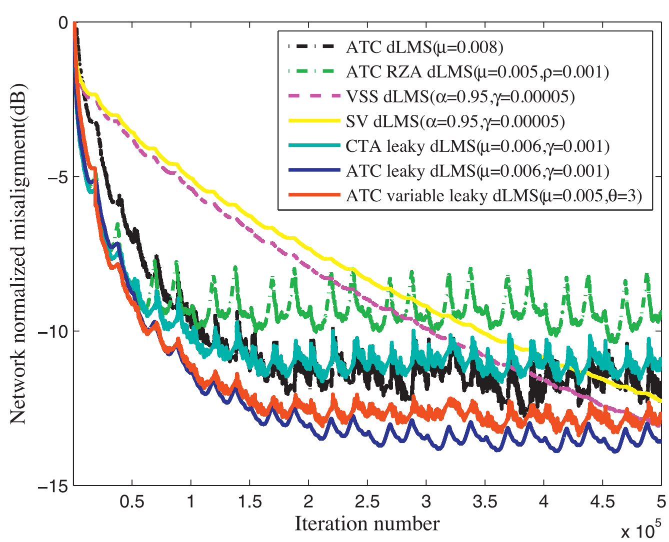

Fig. 10. NMSD of the algorithms for SAEC network.

normalized misalignment was employed:

network normalizd misalignment

$$= 20\log\left(\frac{1}{N}\sum_{k=1}^{N}\mathrm{E}\left\{\frac{\|\boldsymbol{w}_{k,i}-\boldsymbol{h}_k\|_2}{\|\boldsymbol{h}_k\|_2}\right\}\right). \quad (60)$$

The network normalized misalignment of the proposed algorithms was compared with that of the ATC dLMS, VSS dLMS, SV dLMS and ATC RZA dLMS algorithms. The performance of the diffusion algorithms when the speech signal of the SAEC network is presented in Fig. 10 (one trial). Observe that the ATC RZA dLMS algorithm is more sensitive to high background noise than the other algorithms, and it achieves higher steady state error. The VSS dLMS and SV dLMS algorithms, have small misadjustment for SAEC network, but suffer from slow convergence speed. The learning rate of the ATC dLMS algorithm is slower than the ATC RZA dLMS algorithm, and it achieves lower misalignment than the zero attraction strategy. The proposed algorithms achieve the same convergence rate in this case. Particularly, the ATC-based algorithms have smaller lower misalignment compared with the CTA leaky dLMS algorithm. Among these algorithms, the ATC leaky dLMS algorithm holds the best performance for SAEC network.

The proposed algorithms enhance the performance of the existing algorithms for SAEC systems. In practical areas, the sparse echo path is almost universal. The proposed algorithms are not to exploit the sparseness of impulse responses, but still achieve significantly faster adaptation than the conventional dLMS algorithms. Particularly, they even outperform the RZA dLMS algorithm which is based on the l_1-norm. There are several sparsification methods available for AEC systems, such as l_0-norm based algorithms [61,62], proportionate-based algorithms [10,63,64]. In future study, we will attempt to investigate whether these methods can be adopted to develop an efficient diffusion algorithm for SAEC network, and we will consider the logarithmic cost function to improve the performance in impulsive noise environments [65,66].

6. Conclusion

In this paper, the leaky dLMS algorithm with ATC and CTA strategies has been proposed. The proposed algorithm updates the local estimation based on leaky method, which not only maintains the stability for distributed estimation, but also accelerates convergence rate for echo cancellation. Moreover, we have derived and verified the stochastic behavior and stability of the proposed leaky dLMS algorithm. We further proposed the variable leaky dLMS algorithm, which exhibits close performance and does not need the leaky factor selection with the moderate computational complexity compared with the leaky dLMS algorithm. Simulation results in the context of system identification and SAEC network have shown that the proposed algorithms are superior to the competing algorithms.

Acknowledgments

The authors want to express their deep thanks to the anonymous reviewers for many valuable comments which greatly helped to improve the quality of this work.

The work partially supported by the National Science Foundation of P.R. China (Grant: 61571374, 61271340, 61433011). The first author would also like to acknowledge the China Scholarship Council (CSC) for providing him with financial support to study abroad (No. 201607000050).

Appendix A. Algorithms for SAEC network

Fig. 11 depicts the widely linear model for SAEC, where two input (or loudspeaker) signals at kth node and time i are denoted by $\boldsymbol{u}_{k,1,i}$ and $\boldsymbol{u}_{k,2,i}$, and two output (or microphone) signals denoted by $d_{k,1}(i)$ and $d_{k,2}(i)$. In the near-end, the microphone signals are obtained as

$$d_{k,1}(i) = y_{k,1}(i) + v_{k,1}(i) \tag{61}$$

$$d_{k,2}(i) = y_{k,2}(i) + v_{k,2}(i) \tag{62}$$

where $y_{k,1}(i)$ and $y_{k,2}(i)$ denote the stereo echo signals at agent k, and $v_{k,1}(i)$ and $v_{k,2}(i)$ are background noise signals at agent k. The echo signals can be expressed as

$$y_{k,1}(i) = \boldsymbol{w}_{k,11}^T \boldsymbol{u}_{k,1,i} + \boldsymbol{w}_{k,21}^T \boldsymbol{u}_{k,2,i} \tag{63}$$

$$y_{k,2}(i) = \boldsymbol{w}_{k,12}^T \boldsymbol{u}_{k,1,i} + \boldsymbol{w}_{k,22}^T \boldsymbol{u}_{k,2,i} \tag{64}$$

where $\boldsymbol{w}_{k,11}$, $\boldsymbol{w}_{k,12}$, $\boldsymbol{w}_{k,21}$, $\boldsymbol{w}_{k,22}$ are M-dimensional vectors of the loudspeaker-to-microphone (true) acoustic impulse responses, and $\boldsymbol{u}_{k,1,i} = [u_{k,1,i}, u_{k,1,i-1}, \ldots, u_{k,1,i-M+1}]^T$ and $\boldsymbol{u}_{k,2,i} = [u_{k,2,i}, u_{k,2,i-1}, \ldots, u_{k,2,i-M+1}]^T$ are loudspeaker signal vectors.

Using the widely linear (WL) model [67,68], the two-input/two-output system with real random variables are converted to a single-input/single-output system with complex random variables. Here, we form the complex output signal as

$$d_k(i) = d_{k,1}(i) + j d_{k,2}(i) = y_k(i) + v_k(i) \tag{65}$$

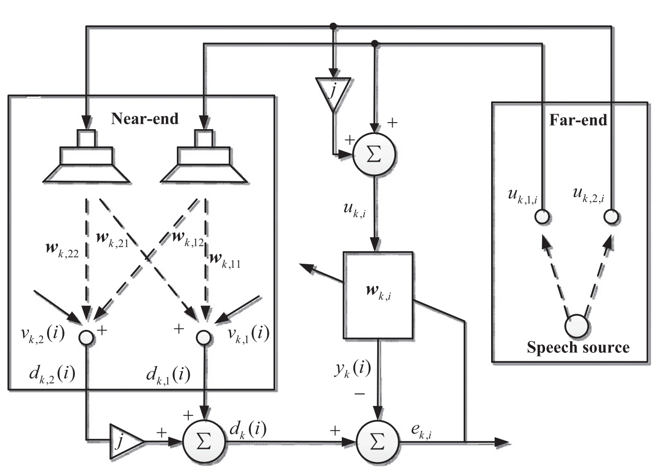

Fig. 11. Widely linear model for SAEC network at node k.

where $j = \sqrt{-1}$, $y_k(i) = y_{k,1}(i) + j y_{k,2}(i)$, and $v_k(i) = v_{k,1}(i) + j v_{k,2}(i)$.

Also, let us define the complex input vector

$$\boldsymbol{u}_{k,i} = \boldsymbol{u}_{k,1,i} + j \boldsymbol{u}_{k,2,i}. \tag{66}$$

Consequently, the complex echo signal can be obtained as

$$y_k(i) = \boldsymbol{w}_{k,t}^H \boldsymbol{u}_{k,i} + \boldsymbol{w}_{k,t}^{*\ H} \boldsymbol{u}_{k,i}^* \tag{67}$$

where the superscripts H and $*$ denote the transpose-conjugate and conjugate, respectively, and

$$\boldsymbol{w}_{k,t} = \boldsymbol{w}_{k,t1} + j \boldsymbol{w}_{k,t2} \tag{68}$$

$$\boldsymbol{w}'_{k,t} = \boldsymbol{w}'_{k,t1} + j \boldsymbol{w}'_{k,t2} \tag{69}$$

with

$$\boldsymbol{w}_{k,t1} = \frac{\boldsymbol{w}_{k,11} + \boldsymbol{w}_{k,22}}{2}, \boldsymbol{w}_{k,t2} = \frac{\boldsymbol{w}_{k,21} - \boldsymbol{w}_{k,12}}{2}, \tag{70}$$

$$\boldsymbol{w}'_{k,t1} = \frac{\boldsymbol{w}_{k,11} - \boldsymbol{w}_{k,22}}{2}, \boldsymbol{w}'_{k,t2} = -\frac{\boldsymbol{w}_{k,21} + \boldsymbol{w}_{k,12}}{2}. \tag{71}$$

Then, the complex output signal can be rewritten as

$$d_k(i) = \boldsymbol{h}_k^H \boldsymbol{u}_{k,i} + v_k(i) \tag{72}$$

where $\boldsymbol{h}_k = [\boldsymbol{w}_{k,t}^T, \boldsymbol{w}'^{\,T}_{k,t}]^T$ is the complex acoustic impulse response of length $2M$, and $\boldsymbol{u}_{k,i} = [\boldsymbol{u}_{k,i}^T, \boldsymbol{u}_{k,i}^H]^T$.

Since the two loudspeaker (input) signals are linearly related which results in a non-unique solution problem, it is necessary to preprocess the input signals to weaken their coherence. A simple but efficient nonlinear method is to use positive and negative half-wave rectifiers at each channel [67,68]

$$u'_{k,1,i} = u_{k,1,i} + b\frac{u_{k,1,i} + |u_{k,1,i}|}{2} \tag{73}$$

$$u'_{k,2,i} = u_{k,2,i} + b\frac{u_{k,2,i} - |u_{k,2,i}|}{2} \tag{74}$$

where b is a parameter used to control the amount of nonlinearity.

To implement the ATC leaky dLMS algorithm for SAEC, one should compute the follows at each iteration

ATC leaky dLMS algorithm:

$$\begin{cases} \boldsymbol{\varphi}_{k,i} = (1-\mu\gamma)\boldsymbol{w}_{k,i-1} + \mu \sum\limits_{l\in\mathcal{N}_k} c_{l,k} \boldsymbol{u}_{l,i}^T e_{l,i}^*(i) \\ \boldsymbol{w}_{k,i} = \sum\limits_{l\in\mathcal{N}_k} a_{l,k} \boldsymbol{\varphi}_{l,i} \end{cases} \tag{75}$$

where $\{c_{l,k}\}$ denote the weighting coefficients, which are the real, non-negative, and meet the condition $c_{l,k} = 0$ if $l \notin \mathcal{N}_k$. The diffusion adaptation formulas of the CTA leaky dLMS algorithm for SAEC system can be easily obtained by reversing the step order of (75).

References

[1] J. Chen, C. Richard, J.C.M. Bermudez, P. Honeine, Nonnegative least-mean-square algorithm, IEEE Trans. Signal Process. 59 (11) (2011) 5225–5235.
[2] L. Lu, H. Zhao, K. Li, B. Chen, A novel normalized sign algorithm for system identification under impulsive noise interference, Circuits Syst. Signal Process. 35 (9) (2016) 3244–3265.
[3] B. Huang, Y. Xiao, Y. Ma, G. Wei, J. Sun, A simplified variable step-size LMS algorithm for fourier analysis and its statistical properties, Signal Process. 117 (2015) 69–81.
[4] S. Zhang, J. Zhang, Robust bounding ellipsoidal adaptive constrained least-squares algorithm and its performance analysis, Digit. Signal Process. 51 (2016) 124–132.
[5] H.-Y. Lin, C.-C. Hu, Y.-F. Chen, J.-H. Wen, An adaptive robust LMS employing fuzzy step size and partial update, IEEE Signal Process. Lett. 12 (8) (2005) 545–548.
[6] K. Mayyas, T. Aboulnasr, Leaky LMS algorithm: MSE analysis for gaussian data, IEEE Trans. Signal Process. 45 (4) (1997) 927–934.

[7] S.C. Douglas, Performance comparison of two implementations of the leaky LMS adaptive filter, IEEE Trans. Signal Process. 45 (8) (1997) 2125–2129.
[8] A.H. Sayed, T.Y. Al-Naffouri, Mean-square analysis of normalized leaky adaptive filters, IEEE ICASSP 6 (2001) pp.3873–3876.
[9] B.D. Rigling, Subspace leaky LMS, IEEE Signal Process. Lett. 11 (2) (2004) 136–139.
[10] G.B. Chandra, A. Mitra, A fast adaptive echo canceler with leaky proportionate NLMS algorithm, Information and Communication Technology in Electrical Sciences (ICTES 2007) (2007) pp. 611–615.
[11] O.J. Tobias, R. Seara, Leaky-FXLMS algorithm: stochastic analysis for Gaussian data and secondary path modeling error, IEEE Trans. Speech Audio Process. 13 (6) (2005) 1217–1230.
[12] J. Cheer, S.J. Elliott, Active noise control of a diesel generator in a luxury yacht, Appl. Acoust. 105 (2016) 209–214.
[13] A.M. Al Omour, A. Zidouri, N. Iqbal, A. Zerguine, Filtered-x least mean fourth (FXLMF) and leaky FXLMF adaptive algorithms, EURASIP J. Adv. Signal Process. 2016 (1) (2016) 1–20.
[14] E. Horita, K. Sumiya, H. Urakami, S. Mitsuishi, A leaky RLS algorithm: its optimality and implementation, IEEE Trans. Signal Process. 52 (10) (2004) 2924–2936.
[15] D.B. Bhoyar, C. Dethe, M. Mushrif, A.P. Narkhede, Leaky least mean square (LLMS) algorithm for channel estimation in BPSK-QPSK-PSK MIMO-OFDM system, International Multi-Conference on Automation, Computing, Communication, Control and Compressed Sensing.(2013) 623–629.
[16] J.M. Gil-Cacho, M. Signoretto, T. van Waterschoot, M. Moonen, S.H. Jensen, Nonlinear acoustic echo cancellation based on a sliding-window leaky kernel affine projection algorithm, IEEE Trans. Audio Speech Lang. Process. 21 (9) (2013) 1867–1878.
[17] O.U.R. Khattak, A. Zerguine, Leaky least mean fourth adaptive algorithm, IET Signal Process. 7 (2) (2013) 134–145.
[18] B. Huang, L. Wen, Y. Xiao, J. Sun, Y. Shen, A complete parallel narrow band active noise control system based on residual error separation using variable leaky LMS, Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA) (2015) 981–985.
[19] S.R. Arya, B. Singh, Performance of DSTATCOM using leaky LMS control algorithm, IEEE Trans. Circuits Syst. II 1 (2) (2013) 104–113.
[20] M.S. Salman, Sparse leaky-LMS algorithm for system identification and its convergence analysis, Int. J. Adapt. Control Signal Process. 28 (10) (2014) 1065–1072.
[21] R. Abdolee, B. Champagne, A.H. Sayed, Estimation of space-time varying parameters using a diffusion LMS algorithm, IEEE Trans. Signal Process. 62 (2) (2014) 403–418.
[22] R. Abdolee, S. Saur, B. Champagne, A.H. Sayed, Diffusion LMS localization and tracking algorithm for wireless cellular networks, IEEE International Conference on Acoustics, Speech Signal Processing (2013a) pp. 4598–4602.
[23] R. Abdolee, B. Champagne, A.H. Sayed, Diffusion LMS strategies for parameter estimation over fading wireless channels, IEEE International Conference on Communications (ICC) (2013b) pp. 1926–1930.
[24] R. Abdolee, B. Champagne, A.H. Sayed, Diffusion LMS for source and process estimation in sensor networks, IEEE Statistical Signal Processing Workshop (SSP)(2012) pp. 165–168.
[25] R. Abdolee, B. Champagne, Diffusion LMS algorithms for sensor networks over non-ideal inter-sensor wireless channels, International Conference on Distributed Computing in Sensor Systems and Workshops (DCOSS) (2011) pp. 1–6.
[26] S.-Y. Tu, A.H. Sayed, Distributed decision-making over adaptive networks, IEEE Trans. Signal Process. 62 (5) (2014) 1054–1069.
[27] A.H. Sayed, S.-Y. Tu, J. Chen, X. Zhao, Z.J. Towfic, Diffusion strategies for adaptation and learning over networks: an examination of distributed strategies and network behavior, IEEE Signal Process. Mag. 30 (3) (2013) 155–171.
[28] L. Li, J.A. Chambers, C.G. Lopes, A.H. Sayed, Distributed estimation over an adaptive incremental network based on the affine projection algorithm, IEEE Trans. Signal Process. 58 (1) (2010) 151–164.
[29] M.O.B. Saeed, A.U. Sheikh, A new LMS strategy for sparse estimation in adaptive networks, in: 23rd International Symposium on Personal Indoor and Mobile Radio Communications (PIMRC), IEEE, 2012, pp. 1722–1733.
[30] Y. Liu, W.K.S. Tang, Enhanced incremental LMS with norm constraints for distributed in-network estimation, Signal Process. 94 (2014) 373–385.
[31] M.S.E. Abadi, A.-R. Danaee, Low computational complexity family of affine projection algorithms over adaptive distributed incremental networks, Int. J. Electron. Commun. 68 (2) (2014) 97–110.
[32] R. Abdolee, V. Vakilian, B. Champagne, Tracking performance and optimal step-sizes of diffusion LMS algorithms in nonstationary signal environment, IEEE Trans. Control Network Syst. (2016).
[33] J. Chen, A.H. Sayed, Diffusion adaptation strategies for distributed optimization and learning over networks, IEEE Trans. Signal Process. 60 (8) (2012) 4289–4305.
[34] R. Abdolee, B. Champagne, A.H. Sayed, Diffusion adaptation over multi-agent networks with wireless link impairments, IEEE Trans. Mobile Comput. 15 (6) (2016) 1362–1376.
[35] R. Abdolee, B. Champagne, Diffusion LMS strategies in sensor networks with noisy input data, IEEE/ACM Trans. Netw. 24 (1) (2016) 3–14.
[36] Y. Liu, C. Li, Z. Zhang, Diffusion sparse least-mean squares over networks, IEEE Trans. Signal Process. 60 (8) (2012) 4480–4485.
[37] F.S. Cattivelli, A.H. Sayed, Diffusion LMS strategies for distributed estimation, IEEE Trans. Signal Process. 58 (3) (2010) 1035–1048.
[38] 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. 2013 (1) (2013a) 135.
[39] 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) (2013b) 827–845.
[40] M.O.B. Saeed, A. Zerguine, M.S. Sohail, S. Rehman, W. Ejaz, A. Anpalagan, Variable step-size strategy for distributed parameter estimation of compressible systems in WSNs, in: Computer Aided Modelling and Design of Communication Links and Networks (CAMAD), 2016 IEEE 21st International Workshop on, IEEE, 2016, pp. 1–5.
[41] D.-C. Ahn, J.-W. Lee, S.-J. Shin, W.-J. Song, A new robust variable weighting coefficients diffusion LMS algorithm, Signal Process. 131 (2017) 300–306.
[42] R. Arablouei, S. Werner, Y.-F. Huang, K. Doğançay, Distributed least mean-square estimation with partial diffusion, IEEE Trans. Signal Process. 62 (2) (2014a) 472–484.
[43] R. Arablouei, K. Doğançay, S. Werner, Y.-F. Huang, Adaptive distributed estimation based on recursive least-squares and partial diffusion, IEEE Trans. Signal Process. 62 (14) (2014b) 3510–3522.
[44] R. Arablouei, S. Werner, K. Doğançay, Partial-diffusion recursive least-squares estimation over adaptive networks, in: 5th International Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP), IEEE, 2013, pp. 89–92.
[45] Ø.L. Rørtveit, J.H. Husøy, A.H. Sayed, Diffusion LMS with communication constraints, in: Conference Record of the Forty Fourth Asilomar Conference on Signals, Systems and Computers (ASILOMAR), IEEE, 2010, pp. 1645–1649.
[46] S. Chouvardas, K. Slavakis, S. Theodoridis, Trading off complexity with communication costs in distributed adaptive learning via Krylov subspaces for dimensionality reduction, IEEE J. Selected Topics Signal Process. 7 (2) (2013) 257–273.
[47] M.O. Sayin, S.S. Kozat, Compressive diffusion strategies over distributed networks for reduced communication load, IEEE Trans. Signal Process. 62 (20) (2014) 5308–5323.
[48] R. Arablouei, S. Werner, K. Doğançay, Y.-F. Huang, Analysis of a reduced-communication diffusion LMS algorithm, Signal Process. 117 (2015) 355–361.
[49] J. Ni, Diffusion sign subband adaptive filtering algorithm for distributed estimation, IEEE Signal Process. Lett. 22 (11) (2015) 2029–2033.
[50] C. Li, P. Shen, Y. Liu, Z. Zhang, Diffusion information theoretic learning for distributed estimation over network, IEEE Trans. Signal Process. 61 (16) (2013) 4011–4024.
[51] W. Ma, B. Chen, J. Duan, H. Zhao, Diffusion maximum correntropy criterion algorithms for robust distributed estimation, Digit. Signal Process. 58 (7) (2016) 10–19.
[52] M. Sowjanya, A. Sahoo, S. Kumar, Distributed incremental leaky LMS, International Conference on Communications and Signal Processing (ICCSP) (2015) pp. 1753–1757.
[53] C.G. Lopes, A.H. Sayed, Randomized incremental protocols over adaptive networks, 2010 IEEE International Conference on Acoustics, Speech and Signal Processing (2010) pp.3514–3517.
[54] B. Subudhi, P. Ray, S. Ghosh, Variable leaky least mean-square algorithm-based power system frequency estimation, IET Sci. Meas. Technol. 6 (4) (2012) 288–297.
[55] O.J. Tobias, R. Seara, On the LMS algorithm with constant and variable leakage factor in a nonlinear environment, IEEE Trans. Signal Process. 54 (9) (2006) 3448–3458.
[56] S. Jo, J. Choi, Y.H. Lee, Modified leaky LMS algorithm for channel estimation in DS-CDMA system, IEEE Commun. Lett. 6 (5) (2002) 202–204.
[57] A. Sayed, Fundamentals of adaptive filtering, Wiley, New York, 2003.
[58] P.D. Lorenzo, A.H. Sayed, Sparse distributed learning based on diffusion adaptation, IEEE Trans. Signal Process. 61 (6) (2013) 1419–1433.
[59] V.D. Blondel, J.M. Hendrickx, A. Olshevsky, J.N. Tsitsiklis, Convergence in multiagent coordination, consensus, and flocking, in Proc. Joint 44th IEEE Conf. Decision Control Eur. Control Conf.(CDC-ECC) (2005) pp.2996–3000..
[60] H. Zhao, Y. Yu, S. Gao, X. Zeng, Z. He, Memory proportionate APA with individual activation factors for acoustic echo cancellation, IEEE Trans. Audio, Speech, Lang. Process. 22 (6) (2014) 1047–1055.
[61] Y. Gu, J. Jin, S. Mei, l_0 Norm constraint LMS algorithm for sparse system identification, IEEE Signal Process. Lett. 16 (9) (2009) 774–777.
[62] M.V. Lima, T.N. Ferreira, W.A. Martins, P.S. Diniz, Sparsity-aware data-selective adaptive filters, IEEE Trans. Signal Process. 62 (17) (2014) 4557–4572.
[63] L. Lu, H. Zhao, C. Chen, A normalized subband adaptive filter under minimum error entropy criterion, Signal Image Video Process. 10 (6) (2016) 1097–1103.
[64] R.L. Das, M. Chakraborty, On convergence of proportionate-type normalized least mean square algorithms, IEEE Trans. Circuits Syst. II 62 (5) (2015) 491–495.
[65] L. Lu, H. Zhao, B. Chen, Improved-variable-forgetting-factor recursive algorithm based on the logarithmic cost for volterra system identification, IEEE Trans. Circuits Syst. II 63 (6) (2016) 588–592.
[66] L. Lu, H. Zhao, Adaptive volterra filter with continuous l_p-norm using a logarithmic cost for nonlinear active noise control, J. Sound Vibrat. 364 (2016) 14–29.
[67] C. Stanciu, J. Benesty, C. Paleologu, T. Gänsler, S. Ciochină, A widely linear model for stereophonic acoustic echo cancellation, Signal Process. 93 (2013) 511–516.
[68] C. Paleologu, J. Benesty, S. Ciochină, Widely linear general Kalman filter for stereophonic acoustic echo cancellation, Signal Process. 94 (2014) 570–575.