Analysis of Partial Diffusion LMS for Adaptive Estimation Over Networks with Noisy Links

Vahid Vahidpour, Amir Rastegarnia, Azam Khalili, Wael M. Bazzi, and Saeid Sanei, *Senior Member, IEEE*

***Abstract*—In partial diffusion-based least mean square (PDLMS) scheme, each node shares a part of its intermediate estimate vector with its neighbors at each iteration. In this paper, besides being involved in more general PDLMS scheme, we figure out how the noisy links affect deterioration of network performance during the exchange of weight estimates. We investigate the steady state mean square deviation (MSD) and derive a theoretical expression for it. We demonstrate that the PDLMS algorithm is stable and convergent in both mean and mean-square sense under non-ideal links. However, unlike the established statements on PDLMS scheme under ideal links, the trade-off between MSD performance and the number of selected entries of the intermediate estimate vectors as a sign of communication cost is mitigated. Strictly speaking, considering non-ideal links condition adds a new complexity to MSD relation that has a noticeable effect on its performance. This term violates the trade-off between communication cost and estimation performance of the networks in comparison to noise-free condition on the links. Our simulation results substantiate the effect of noisy links on PDLMS algorithm and verify the theoretical findings. They match well with theory.**

***Index Terms*—Adaptive networks, distributed estimation, least mean-square, noisy links, partial diffusion.**

I. Introduction

WE consider the problem of distributed estimation in the diffusion adaptive networks context, where the spatially-scattered nodes have adaptation and learning capabilities. In such networks, the nodes are linked together through a topology and exchange information through localized in-network processing to perform decentralized information processing and optimization in a cooperative and online manner. The local interactions and diffusion of information across the network enable the nodes to respond in real-time to the drifts in statistical properties of the data and to the changes in network topology [1]–[4]. Several strategies for distributed estimation over adaptive networks have been reported in the literature. Diffusion strategies [5]–[10] are among the most popular propositions, in the literature. They are scalable as well as robust to link/node failure and have good adaptability and tracking performance with respect to other strategies [2]. In adaptive diffusion implementations, the nodes communicate with their immediate neighbors and the information is processed locally and simultaneously at all nodes across the network.

It is obvious that the weight estimates that are exchanged among the nodes can be subject to different perturbations such as quantization errors, noisy input data, additive noise over the communication links and wireless link impairments. Studying the degradation in performance that results from the mentioned perturbations can be found in [11]–[17].

Due to the limited power and bandwidth resources for communication among nodes over a distributed networks (such as wireless sensor networks), the most expensive part of realizing a cooperative task is the data transmission through the links. Generally speaking, although the benefits of diffusion strategies achieved by increasing internode communications, they compromised by the communication cost. As the consequence, since different nodes can have different numbers of neighbors, they may require disparate hardware or consume power differently. Therefore, reducing the amount of internode communications, while maintaining the benefits of cooperation is of practical importance. There have been several efforts to achieve the mentioned objective such as reducing the dimension of the estimates [18]–[20], selecting a subset of the entries of the intermediate estimate vectors [21], [22], set-membership filtering [23]–[25] and partial updating [26].

Among these methods, we focus on [21] where the LMS algorithm for adaptive distribute estimation has been formulated and analyzed by utilizing partial-diffusion. In [21], an adapt-then-combine (ATC) PDLMS algorithm has been reported for distributed estimation over adaptive networks with ideal links. In the mentioned algorithm, at each iteration, each node transmits a subset of the entries of intermediate estimate vector to its neighbors. However, as we mentioned earlier, in practice the weight estimates that are exchanged among the nodes can be subject to additive noise over communication links. In this paper, besides being involved in more general PDLMS scheme, we figure out how the noisy links affect deterioration of network performance during the exchange of weight estimates. Among other results, our analysis provides some useful insights on the communication cost and estimation performance trade-off for general PDLMS scheme under non-ideal links. Our main contributions in this paper include:

(i) Focusing on [21] which involves transmission of a subset of the entries of the internode estimate vectors named partial diffusion, we provide a more general algorithmic structure of which [21] is just a special case. To achieve this, we consider the fact that weight estimates exchanged among the nodes can be subject to quantization errors and additive noise over communication links. Like

V. Vahidpour, A. Rastegarnia, and A. Khalili and are with the Department of Electrical Engineering, Malayer University, Malayer 65719-95863, Iran (email: v.vahidpour@gmail.com; rastegar@tabrizu.ac.ir; a-khalili@tabrizu.ac.ir).

W. M. Bazzi is with the Department of Electrical Engineering, American University in Dubai, Dubai, United Arab Emirates, (email: wbazzi@and.edu)

S. Sanei is with the Department of Computer Science, University of Surrey, Surrey GU2 7XH, UK (email: s.sanei@surrey.ac.uk).

[21], we also consider two different schemes for selecting the weight vector entries for transmission at each iteration. We allow for noisy exchange just during the two combination steps. It should be noted that since our objective is to minimize the internode communication, the nodes only exchange their intermediate estimates with their neighbors;

(ii) Using the energy conservation argument [27] we analyze the stability of algorithms in mean and mean square senses under certain statistical conditions.

(iii) We illustrate the comparable convergence performance of PDLMS algorithm with noisy links in different numerical examples.

The main aim of this paper is that the noisy links are the main factor in performance degradation of a partial diffusion least mean squares (PLMS) algorithm running in a network with noisy links. In other words, considering noisy links adds an extra term to MSD relation. This term seriously upset the balance of the trading off between communication cost and the estimation performance, in comparison with the ideal case. Because, the more entries are communicated at each iteration, the more perturbed weight estimates are interred in the consultation phase.

This work is organized as follows. In Section II, we formulate the PDLMS under noisy information exchange. The performance analyses are examined in Section III. We provide simulation results in Section IV and draw the conclusions in Section V.

A. Notation

We use the lowercase letters to denote vectors, uppercase letter for matrices, plain letter for deterministic variables, and the boldface letters for random variables. We also use $(\cdot)^*$ to denote conjugate transposition, $\operatorname{tr}(\cdot)$ for the trace of matrix, $\otimes$ for Kronecker product, and $\operatorname{vec}\{\cdot\}$ for a vector formed by stacking the columns of its matrix argument. We further use $\operatorname{diag}\{\cdot\}$ to denote a (block) diagonal matrix formed from its argument, and $\operatorname{col}\{\cdot\}$ to denote a column vector formed by stacking its arguments on top of each other. All vectors in our treatment are column vectors, with the exception of regression vectors, $\boldsymbol{u}_{k,i}$.

II. Partial Diffusion Algorithms with Noisy Information Exchange

Consider a connected network consisting of N nodes. Each node k collects scalar measurements $\boldsymbol{d}_{k,i}$ and $1\times M$ regression data vectors $\boldsymbol{u}_{k,i}$ over successive time instants $i \geq 0$. Note that we use parenthesis to refer to the time-dependence of scalar variables, as in $\boldsymbol{d}_{k,i}$, and subscripts to refer to the time-dependence of vector variables, as in $\boldsymbol{u}_{k,i}$. The measurements across all nodes are assumed to be related to an unknown $M\times 1$ vector w^o via linear regression model of the form [27]:

$$\boldsymbol{d}_{k,i} = \boldsymbol{u}_{k,i} w^o + \boldsymbol{v}_{k,i} \tag{1}$$

where $\boldsymbol{v}_{k,i}$ denotes the measurement or model noise. We are now interested in solving optimization problems of the type:

$$w^o = \min_{w} \sum_{k=1}^{N} \mathbb{E}\left[|\boldsymbol{d}_{k,i} - \boldsymbol{u}_{k,i}\boldsymbol{w}|^2 \right] \tag{2}$$

The nodes in the network would like to cooperate with each other in order to estimate w^o by solving the equation above in an adaptive manner. Putting an accurate interpretation on solution vector w^o from (2) depends on application under consideration. One possible interpretation is that the entries of w^o represent the location coordinates of a flying object (such as tracking a projectile) that agents are trying to find. In other applications, the entries of w^o describes an underlying tapped-delay-line model also known as finite-impulse-response (FIR) that agents are interested in estimating the parameters of an FIR model, such as taps of a communication channel or the parameters of some (approximate) model of interest in finance or biology [2]. We review the diffusion adaptation strategies with noisy links below.

A. Diffusion Adaptation with Noisy Information Exchange

Consider the following general adaptive diffusion strategies corresponding to the case in which the nodes only share weight estimates for $i \geq 0$:

$$\boldsymbol{\phi}_{k,i} = \sum_{l\in\mathcal{N}_k} c_{1,lk} \boldsymbol{w}_{l,i-1} \tag{3}$$

$$\boldsymbol{\psi}_{k,i} = \boldsymbol{\phi}_{k,i-1} + \mu_k \boldsymbol{u}^*_{k,i}\left[\boldsymbol{d}_{k,i} - \boldsymbol{u}_{k,i}\boldsymbol{\phi}_{k,i-1}\right] \tag{4}$$

$$\boldsymbol{w}_{k,i} = \sum_{l\in\mathcal{N}_k} c_{2,lk} \boldsymbol{\psi}_{l,i} \tag{5}$$

Local estimators of w^o, that node k computes based on observations $\{\boldsymbol{d}_k(j), \boldsymbol{u}_{k,j} | j \leq i\}$ in addition to intermediate estimators up to and including time i, are denoted by $M\times 1$ vectors $\left\{\boldsymbol{\phi}_{k,i}, \boldsymbol{\psi}_{k,i}, \boldsymbol{w}_{k,i}\right\}$. The physical meanings of these vectors and w^o are exactly the same. The scalars $\{c_{1,lk}, c_{2,lk}\}$ are non-negative real coefficients corresponding to the (l,k) entries of $N\times N$ combination matrices $\{C_1, C_2\}$, respectively. They are zero whenever node $l \notin \mathcal{N}_k$, where $\mathcal{N}_k$ denotes the neighborhood of node k. These matrices are assumed to satisfy the conditions:

$$C_1^T = 1_N,\ C_2^T = 1_N \tag{6}$$

where the notation 1_N denotes an $N\times 1$ column vector with all its entries equal to one.

We model the noisy data received by node k from its neighbor l as follows:

$$\boldsymbol{w}_{lk,i-1} = \boldsymbol{w}_{l,i-1} + \boldsymbol{v}^{(w)}_{lk,i-1} \tag{7}$$

$$\boldsymbol{\psi}_{lk,i} = \boldsymbol{\psi}_{l,i} + \boldsymbol{v}^{(\psi)}_{lk,i} \tag{8}$$

where $\boldsymbol{v}^{(w)}_{lk,i-1}$ $(M\times 1)$ and $\boldsymbol{v}^{(\psi)}_{lk,i}$ $(M\times 1)$ are the noise observations. It should be noted that the subscript lk indicates that l is the source and k is the sink and the flow of information is from l to k.

Using the perturbed data (7) and (8), the adaptive strategy (3)-(5) becomes

$$\phi_{k,i} = \sum_{l\in\mathcal{N}_k} c_{1,lk}\boldsymbol{w}_{lk,i-1} \tag{9}$$

$$\psi_{k,i} = \phi_{k,i-1} + \mu_k \boldsymbol{u}^*_{k,i}\left[\boldsymbol{d}_{k,i} - \boldsymbol{u}_{k,i}\phi_{k,i-1}\right] \tag{10}$$

$$\boldsymbol{w}_{k,i} = \sum_{l\in\mathcal{N}_k} c_{2,lk}\psi_{lk,i} \tag{11}$$

B. Partial Diffusion with Noisy Links

In this paper, we adopt a similar approach proposed in [21] and build up our algorithm upon it. Selecting and scattering L out of M, $0 \le L \le M$, entries of the intermediate estimate vector of each node k at time instant i, make the realization of reducing internode communication possible. According to this scheme, the selection of to be scattered elements could be realized by a diagonal selection matrix, $\mathcal{K}_{k,i}$. Multiplication of $\left\{\boldsymbol{w}_{lk,i-1}, \psi_{lk,i}\right\}$ by $\mathcal{K}_{k,i}$ that have L ones and $M-L$ zeros on its diagonal replaces its non-selected entries with zero. The positions of the ones on diagonal of $\mathcal{K}_{k,i}$ determine the entries of node k that are selected to diffused at time i. Note that, the integer L is fixed and pre-specified [21]. According to (9) and (11)

$$\begin{aligned}\phi_{k,i} = c_{1,kk}\boldsymbol{w}_{k,i-1} + \sum_{l\in\mathcal{N}_k\setminus\{k\}} c_{1,lk}[&\mathcal{K}_{l,i-1}\boldsymbol{w}_{lk,i-1} \\ &+ (I_M - \mathcal{K}_{l,i-1})\,\boldsymbol{w}_{lk,i-1}]\end{aligned} \tag{12}$$

$$\psi_{k,i} = \phi_{k,i-1} + \mu_k \boldsymbol{u}^*_{k,i}\left[\boldsymbol{d}_{k,i} - \boldsymbol{u}_{k,i}\phi_{k,i-1}\right] \tag{13}$$

$$\begin{aligned}\boldsymbol{w}_{k,i} = c_{2,kk}\psi_{k,i} + \sum_{l\in\mathcal{N}_k\setminus\{k\}} c_{2,lk}[&\mathcal{K}_{l,i}\psi_{lk,i} \\ &+ (I_M - \mathcal{K}_{l,i})\,\psi_{lk,i-1}]\end{aligned} \tag{14}$$

where I_M is the identity matrix of size $M \times M$.

The most fundamental problem, we are faced with, hinges on ambiguities in non-diffused elements of nodes in combination phase. When intermediate estimate are partially transmitted, the non-communicated entries are not available to take part in this phase. However, each node requires all entries of intermediate estimate vectors of its neighbors for combination. To avoid this ambiguity, nodes can replace the entries of their own intermediate estimates instead of the ones from the neighbors that are not available. substitute

$$(I_M - \mathcal{K}_{l,i-1})\,\boldsymbol{w}_{k,i-1}, \quad \forall l \in \mathcal{N}_k\setminus\{k\} \tag{15}$$

for

$$(I_M - \mathcal{K}_{l,i-1})\,\boldsymbol{w}_{lk,i-1}, \quad \forall l \in \mathcal{N}_k\setminus\{k\} \tag{16}$$

and

$$(I_M - \mathcal{K}_{l,i})\,\psi_{k,i}, \quad \forall l \in \mathcal{N}_k\setminus\{k\} \tag{17}$$

for

$$(I_M - \mathcal{K}_{l,i})\,\psi_{lk,i}, \quad \forall l \in \mathcal{N}_k\setminus\{k\} \tag{18}$$

Based on this approach, we formulate general PDLMS under noisy information exchange as follows:

$$\begin{aligned}\phi_{k,i} = c_{1,kk}\boldsymbol{w}_{k,i-1} + \sum_{l\in\mathcal{N}_k\setminus\{k\}} c_{1,lk}[&\mathcal{K}_{l,i-1}\boldsymbol{w}_{lk,i-1} \\ &+ (I_M - \mathcal{K}_{l,i-1})\,\boldsymbol{w}_{k,i-1}]\end{aligned} \tag{19}$$

$$\psi_{k,i} = \phi_{k,i-1} + \mu_k \boldsymbol{u}^*_{k,i}\left[\boldsymbol{d}_{k,i} - \boldsymbol{u}_{k,i}\phi_{k,i-1}\right] \tag{20}$$

$$\begin{aligned}\boldsymbol{w}_{k,i} = c_{2,kk}\psi_{k,i} + \sum_{l\in\mathcal{N}_k\setminus\{k\}} c_{2,lk}[&\mathcal{K}_{l,i}\psi_{lk,i} \\ &+ (I_M - \mathcal{K}_{l,i})\,\psi_{k,i-1}]\end{aligned} \tag{21}$$

Remark. The probability of transmission for all the entries at each node is equal and expressed as

$$\rho = L/M \tag{22}$$

Moreover, the entry selection matrices, $\mathcal{K}_{k,i}$, do not depend on any data/parameter other than L and M.

From (7), (8), expression (19)-(21) can be written as:

$$\begin{aligned}\phi_{k,i-1} = c_{1,kk}\boldsymbol{w}_{k,i-1} + \sum_{l\in\mathcal{N}_k\setminus\{k\}} c_{1,lk}[&\mathcal{K}_{l,i-1}\boldsymbol{w}_{l,i-1} \\ &+ (I_M - \mathcal{K}_{l,i-1})\,\boldsymbol{w}_{k,i-1}] \\ &+ \sum_{l\in\mathcal{N}_k\setminus\{k\}} c_{1,lk}\mathcal{K}_{l,i-1}\boldsymbol{v}^{(w)}_{lk,i-1}\end{aligned} \tag{23}$$

$$\psi_{k,i} = \phi_{k,i-1} + \mu_k \boldsymbol{u}^*_{k,i}\left[\boldsymbol{d}_{k,i} - \boldsymbol{u}_{k,i}\phi_{k,i-1}\right] \tag{24}$$

$$\begin{aligned}\boldsymbol{w}_{k,i} = c_{2,kk}\psi_{k,i} + \sum_{l\in\mathcal{N}_k\setminus\{k\}} c_{2,lk}[&\mathcal{K}_{l,i}\psi_{lk,i} \\ &+ (I_M - \mathcal{K}_{l,i})\,\psi_{k,i}] \\ &+ \sum_{l\in\mathcal{N}_k\setminus\{k\}} c_{2,lk}\mathcal{K}_{l,i}\boldsymbol{v}^{(\psi)}_{lk,i}\end{aligned} \tag{25}$$

Introducing the following aggregate $M \times 1$ zero mean noise signals:

$$\boldsymbol{v}^{(w)}_{k,i-1} = \sum_{l\in\mathcal{N}_k\setminus\{k\}} c_{1,lk}\mathcal{K}_{l,i-1}\boldsymbol{v}^{(w)}_{lk,i-1} \tag{26}$$

$$\boldsymbol{v}^{(\psi)}_{k,i} = \sum_{l\in\mathcal{N}_k\setminus\{k\}} c_{2,lk}\mathcal{K}_{l,i}\boldsymbol{v}^{(\psi)}_{lk,i} \tag{27}$$

where $\left\{\boldsymbol{v}^{(w)}_{k,i-1}, \boldsymbol{v}^{(\psi)}_{k,i}\right\}$ represent the aggregate effect on node k of all selected exchange noises from its neighbors while exchanging the estimates $\left\{\boldsymbol{w}_{l,i-1}, \psi_{l,i}\right\}$ during the two combination steps. The $M \times M$ covariance matrices of these noises are given by:

$$R^{(w)}_{v,k} = \sum_{l\in\mathcal{N}_k\setminus\{k\}} c^2_{1,lk}\rho^2 R^{(w)}_{v,lk} \tag{28}$$

$$R^{(\psi)}_{v,k} = \sum_{l\in\mathcal{N}_k\setminus\{k\}} c^2_{2,lk}\rho^2 R^{(\psi)}_{v,lk} \tag{29}$$

C. Entry Selection Methods

To select L-subset of a set on M elements containing exactly L elements, we employ a similar approach proposed in [21]. Doing so, there exist two different scheme named sequential and stochastic partial-diffusion. These methods are analogous to the selection processes in sequential and stochastic partial-update schemes [26], [28]–[30]. In sequential partial-diffusion the entry selection matrices, $\mathcal{K}_{k,i}$, is diagonal matrix:

$$\mathcal{K}_{k,i} = \begin{bmatrix} \kappa_{1,i} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \kappa_{M,i} \end{bmatrix}, \kappa_{\ell,i} = \begin{cases} 1 & \text{if } \ell \in \mathcal{J}_{(i \bmod \bar{B})+1} \\ 0 & \text{otherwise} \end{cases} \tag{30}$$

with $\bar{B} = \lceil M/L \rceil$. The number of selection entries at each iteration is limited by L. The coefficient subsets $\mathcal{J}_i$ are not unique as long as they meet the following requirements [26]:

1) Cardinality of $\mathcal{J}_i$ is between 1 and L;
2) $\bigcup_{\tau=1}^{\bar{B}} = \mathcal{S}$ where $\mathcal{S} = \{1, 2, \ldots, M\}$;
3) $\mathcal{J}_\tau \cap \mathcal{J}_\eta = \emptyset$, $\forall \tau, \eta \in \{1, \ldots, \bar{B}\}$ and $\tau \neq \eta$.

The description of entry selection matrices, $\mathcal{K}_{k,i}$, in stochastic partial-diffusion approach is similar to that of sequential one. The only difference is as follows. At a given iteration i, the sequential case one of the set $\mathcal{J}_\tau$, $\tau = \{1, \ldots, \bar{B}\}$ is chosen in a predetermined fashion, whereas for stochastic case, one of the sets $\mathcal{J}_\tau$ is sampled at random from $\{\mathcal{J}_1, \mathcal{J}_2, \ldots, \mathcal{J}_{\bar{B}}\}$. One might ask why these methods are considered to organize mentioned selection matrices. To answer this question, it is worth mentioning that the nodes need to know which entries of their neighbors' intermediate estimates have been transmitted at each iteration. These schemes are not subject to such requirements.

III. Performance Analysis

We now move on to examine the behavior of the general PDLMS implementations (23)-(25), and the influence of the mentioned perturbations on convergence and steady-state performance. For this reason, we shall study the convergence of the weight estimates both in the mean and mean-square senses.

Assumptions. In order to make the analysis tractable, we consider the following assumptions on statistical properties of the measurement data and noise signals.

(i) The regression data $\boldsymbol{u}_{k,i}$ are temporally white and spatially independent random variables with zero mean and covariance matrix $R_{u,k} \triangleq \mathbb{E}\left[\boldsymbol{u}_{k,i}^* \boldsymbol{u}_{k,i}\right] \geq 0$ where $k = \{1, \ldots, N\}$.

(ii) The noise signals $\boldsymbol{v}_{k,i}$, $\boldsymbol{v}_{k,i-1}^{(w)}$ and $\boldsymbol{v}_{k,i}^{(\psi)}$ are temporally white and spatially independent random variables with zero mean and co(variance)s $\sigma_{v,k}^2$, $R_{v,k}^{(w)}$ and $R_{v,k}^{(\psi)}$, respectively. In addition, The quantities $\left\{R_{v,lk}^{(w)}, R_{v,lk}^{(\psi)}\right\}$ are all zero if $l \in \mathcal{N}_k$ or when $l = k$.

(iii) The regression data $\{\boldsymbol{u}_{m,i_1}\}$, the model noise signal $\boldsymbol{v}_n(i_2)$, and the link noise signals $\boldsymbol{v}_{l_1 k_1, j_1}^{(w)}$ and $\boldsymbol{v}_{l_2 k_2, j_2}^{(\psi)}$ are mutually independent random variables for all indexes $\{i_1, i_2, j_1, j_2, k_1, k_2, l_1, l_2, m, n\}$.

(iv) The step-sizes, μ_k ,$\forall k$, are small enough such that their squared values are negligible.

We are interested in examining the evolution of the weight-error vectors. To do so, we introduce the error vectors:

$$\tilde{\boldsymbol{\phi}}_{k,i} \triangleq w^o - \boldsymbol{\phi}_{k,i} \tag{31}$$

$$\tilde{\boldsymbol{\psi}}_{k,i} \triangleq w^o - \boldsymbol{\psi}_{k,i} \tag{32}$$

$$\tilde{\boldsymbol{w}}_{k,i} \triangleq w^o - \boldsymbol{w}_{k,i} \tag{33}$$

Substituting the linear model (1) into adaptation step (24) and subtraction of both sides from w^o give:

$$\tilde{\boldsymbol{\psi}}_{k,i} = \left(I_M - \mu_k \boldsymbol{u}_{k,i}^* \boldsymbol{u}_{k,i}\right) \tilde{\boldsymbol{\phi}}_{k,i-1} - \mu_k \boldsymbol{u}_{k,i}^* \boldsymbol{v}_{k,i} \tag{34}$$

Using conditions (6), we can rewrite (23) and (25) as

$$\begin{aligned} \boldsymbol{\phi}_{k,i-1} = &\left(I_M - \sum_{l \in \mathcal{N}_k \setminus \{k\}} c_{1,lk} \mathcal{K}_{l,i-1}\right) \boldsymbol{w}_{k,i-1} \\ &+ \sum_{l \in \mathcal{N}_k \setminus \{k\}} c_{1,lk} \mathcal{K}_{l,i-1} \boldsymbol{w}_{l,i-1} \\ &+ \boldsymbol{v}_{k,i-1}^{(w)} \end{aligned} \tag{35}$$

$$\begin{aligned} \boldsymbol{w}_{k,i} = &\left(I_M - \sum_{l \in \mathcal{N}_k \setminus \{k\}} c_{2,lk} \mathcal{K}_{l,i}\right) \boldsymbol{\psi}_{k,i} \\ &+ \sum_{l \in \mathcal{N}_k \setminus \{k\}} c_{2,lk} \mathcal{K}_{l,i} \boldsymbol{\psi}_{l,i} \\ &+ \boldsymbol{v}_{k,i}^{(\psi)} \end{aligned} \tag{36}$$

Subtracting (35) from

$$\begin{aligned} w^o = &\left(I_M - \sum_{l \in \mathcal{N}_k \setminus \{k\}} c_{1,lk} \mathcal{K}_{l,i-1}\right) w^o \\ &+ \sum_{l \in \mathcal{N}_k \setminus \{k\}} c_{1,lk} \mathcal{K}_{l,i-1} w^o \end{aligned} \tag{37}$$

and (36) from

$$\begin{aligned} w^o = &\left(I_M - \sum_{l \in \mathcal{N}_k \setminus \{k\}} c_{2,lk} \mathcal{K}_{l,i}\right) w^o \\ &+ \sum_{l \in \mathcal{N}_k \setminus \{k\}} c_{2,lk} \mathcal{K}_{l,i} w^o \end{aligned} \tag{38}$$

gives

$$\begin{aligned} \tilde{\boldsymbol{\phi}}_{k,i-1} = &\left(I_M - \sum_{l \in \mathcal{N}_k \setminus \{k\}} c_{1,lk} \mathcal{K}_{l,i-1}\right) \tilde{\boldsymbol{w}}_{k,i-1} \\ &+ \sum_{l \in \mathcal{N}_k \setminus \{k\}} c_{1,lk} \mathcal{K}_{l,i-1} \tilde{\boldsymbol{w}}_{l,i-1} \\ &- \boldsymbol{v}_{k,i-1}^{(w)} \end{aligned} \tag{39}$$

$$\tilde{\boldsymbol{w}}_{k,i} = \left(I_M - \sum_{l\in\mathcal{N}_k\backslash\{k\}} c_{2,lk}\boldsymbol{\mathcal{K}}_{l,i}\right)\tilde{\boldsymbol{\psi}}_{k,i} + \sum_{l\in\mathcal{N}_k\backslash\{k\}} c_{2,lk}\boldsymbol{\mathcal{K}}_{l,i}\tilde{\boldsymbol{\psi}}_{l,i} - \boldsymbol{v}_{k,i}^{(\psi)} \tag{40}$$

To describe these relations in more closed form, we collect the information from across the network into block vectors and matrices. Stacking the error vectors from across all nodes into the following $N\times 1$ block vectors, whose individual entries are of size $M\times 1$ each we have:

$$\tilde{\boldsymbol{\phi}}_i \triangleq \text{col}\left\{\tilde{\boldsymbol{\phi}}_{1,i},\ldots,\tilde{\boldsymbol{\phi}}_{N,i}\right\} \tag{41}$$

$$\tilde{\boldsymbol{\psi}}_i \triangleq \text{col}\left\{\tilde{\boldsymbol{\psi}}_{1,i},\ldots,\tilde{\boldsymbol{\psi}}_{N,i}\right\} \tag{42}$$

$$\tilde{\boldsymbol{w}}_i \triangleq \text{col}\left\{\tilde{\boldsymbol{w}}_{1,i},\ldots,\tilde{\boldsymbol{w}}_{N,i}\right\} \tag{43}$$

Also, collecting the noise signal (26) and (27), and their covariances from across the network into $N\times 1$ block vectors and $N\times N$ block diagonal matrices as follows:

$$\boldsymbol{v}_i^{(w)} \triangleq \text{col}\left\{\boldsymbol{v}_{1,i}^{(w)},\ldots,\boldsymbol{v}_{N,i}^{(w)}\right\} \tag{44}$$

$$\boldsymbol{v}_i^{(\psi)} \triangleq \text{col}\left\{\boldsymbol{v}_{1,i}^{(\psi)},\ldots,\boldsymbol{v}_{N,i}^{(\psi)}\right\} \tag{45}$$

$$R_i^{(w)} \triangleq \text{col}\left\{R_{v,1}^{(w)},\ldots,R_{v,N}^{(w)}\right\} \tag{46}$$

$$R_i^{(\psi)} \triangleq \text{col}\left\{R_{v,1}^{(\psi)},\ldots,R_{v,N}^{(\psi)}\right\} \tag{47}$$

Subsequently, we can verify that

$$\tilde{\boldsymbol{\phi}}_{i-1} = \boldsymbol{\mathcal{A}}_{1,i-1}\tilde{\boldsymbol{w}}_{i-1} - \boldsymbol{v}_{i-1}^{(w)} \tag{48}$$

$$\tilde{\boldsymbol{\psi}}_i = (I_{NM} - \mathcal{M}\boldsymbol{\mathcal{R}}_{u,i}) - \mathcal{M}\boldsymbol{s}_i \tag{49}$$

$$\tilde{\boldsymbol{w}}_i = \boldsymbol{\mathcal{A}}_{2,i}\tilde{\boldsymbol{\psi}}_{i-1} - \boldsymbol{v}_i^{(\psi)} \tag{50}$$

where

$$\mathcal{M} \triangleq \text{diag}\left\{\mu_1 I_M,\ldots,\mu_N I_M\right\} \tag{51}$$

$$\boldsymbol{\mathcal{R}}_{u,i} \triangleq \text{diag}\left\{\boldsymbol{u}_{1,i}^*\boldsymbol{u}_{1,i},\ldots,\boldsymbol{u}_{N,i}^*\boldsymbol{u}_{N,i}\right\} \tag{52}$$

with

$$\mathbb{E}\left[\boldsymbol{\mathcal{R}}_{u,i}\right] = \mathcal{R}_u = \text{diag}\left\{R_{u,1},\ldots,R_{u,N}\right\} \tag{53}$$

$$\boldsymbol{s}_i \triangleq \text{diag}\left\{\boldsymbol{u}_{1,i}^*\boldsymbol{v}_1(i),\ldots,\boldsymbol{u}_{N,i}^*\boldsymbol{v}_{N,i}\right\} \tag{54}$$

Here, $\boldsymbol{s}_i$ denotes $N\times 1$ block column vector, whose entries are of size $M\times 1$ each. Following Assumption (i), we have

$$\mathbb{E}\left[\boldsymbol{s}_i\right] = 0 \tag{55}$$

The covariance matrix of $\boldsymbol{s}_i$ is $N\times N$ block diagonal with blocks of size $M\times M$:

$$S = \mathbb{E}\left[\boldsymbol{s}_i\boldsymbol{s}_i^*\right] = \text{diag}\left\{\sigma_{v,1}^2 R_{u,1},\ldots,\sigma_{v,N}^2 R_{u,N}\right\} \tag{56}$$

I_{NM} is also the identity matrix of size $MN\times MN$. Moreover,

$$\boldsymbol{\mathcal{A}}_{r,i} = \begin{bmatrix} A_{1,1,i} & \cdots & A_{1,N,i} \\ \vdots & \ddots & \vdots \\ A_{N,1,i} & \cdots & A_{N,N,i}\end{bmatrix} \quad ,\forall r\in\{1,2\} \tag{57}$$

where

$$A_{p,q,i} = \begin{cases} I_M - \sum_{l\in\mathcal{N}_p\backslash\{p\}} c_{r,lp}\boldsymbol{\mathcal{K}}_{l,i} & \text{if } p=q \\ c_{r,qp}\boldsymbol{\mathcal{K}}_{q,i} & \text{if } q\in\mathcal{N}_p\backslash\{p\} \\ O_M & \text{otherwise}\end{cases} \tag{58}$$

So that the network weight error vector, $\tilde{\boldsymbol{w}}_i$, evolves according to the following stochastic recursion:

$$\tilde{\boldsymbol{w}}_i = \boldsymbol{\mathcal{A}}_{2,i}\left(I_{NM} - \mathcal{M}\boldsymbol{\mathcal{R}}_{u,i}\right)\boldsymbol{\mathcal{A}}_{1,i-1}\tilde{\boldsymbol{w}}_{i-1} - \boldsymbol{\mathcal{A}}_{2,i}\left(I_{NM} - \mathcal{M}\boldsymbol{\mathcal{R}}_{u,i}\right)\boldsymbol{v}_{i-1}^{(w)} - \boldsymbol{\mathcal{A}}_{2,i}\mathcal{M}\boldsymbol{s}_i - \boldsymbol{v}_i^{(\psi)} \tag{59}$$

A. Convergence in Mean

Taking expectation of both sides of (59) under *Remark* and *Assumptions*, we find that the mean error vector evolves according to the following recursion:

$$\mathbb{E}\left[\tilde{\boldsymbol{w}}_i\right] = Q_2\left(I_{NM} - \mathcal{M}\mathcal{R}_u\right)Q_1\mathbb{E}\left[\tilde{\boldsymbol{w}}_{i-1}\right] \tag{60}$$

where

$$Q_1 = \mathbb{E}\left[\boldsymbol{\mathcal{A}}_{1,i-1}\right] \tag{61}$$

$$Q_2 = \mathbb{E}\left[\boldsymbol{\mathcal{A}}_{2,i}\right] \tag{62}$$

Like [21], $Q_r, r\in\{1,2\}$ can be obtained for both stochastic and sequential partial-diffusion using the definition of $\boldsymbol{\mathcal{A}}_{r,i}, r\in\{1,2\}$, see (61) and (62). What is most noteworthy here is to find the value of each $Q_r, r\in\{1,2\}$ entries after applying expectation operator. Therefore, we can write

$$\mathbb{E}\left[A_{p,q,i}\right] = \begin{cases} \left(1-\rho+\rho c_{r,lp}\right)I_M & \text{if } p=q \\ \rho c_{r,qp} I_M & \text{if } q\in\mathcal{N}_p\backslash\{p\} \\ O_M & \text{otherwise}\end{cases} \tag{63}$$

All the entries of $Q_r, r\in\{1,2\}$ are real and non-negative and all the rows of $Q_r, r\in\{1,2\}$ add up to unity. This property [21] can be established for both stochastic and sequential partial-diffusion schemes and for any value of L .

Theorem 1 (Convergence in Mean). Consider the problem of optimizing the global cost (2), Pick Q_1 and Q_2 with are real non-negative entries and all their rows add up to unity. Assume each node in the network measures data that satisfy conditions described in *Assumptions*, and run adaptive diffusion algorithm (23)-(25). Assume further that the exchange of the variables $\left\{\boldsymbol{w}_{l,i-1},\boldsymbol{\psi}_{l,i}\right\}$ is subject to additive noise as (7) and (8). Moreover, the regressors and desired signals are assumed not to exchang among the nodes. Then, all estimates $\{\boldsymbol{w}_{k,i}\}$ across the network converge in the mean to optimal solution w^o if the step-size parameters $\{\mu_k\}$ satisfy

$$0 < \mu_k < \frac{2}{\lambda_{max}\left\{R_{u,k}\right\}} \quad ,\forall k \tag{64}$$

Proof: The weight error vectors $\{\tilde{\boldsymbol{w}}_i\}$ converge to zero if, and only if, the matrix $Q_2\left(I_{NM} - \mathcal{M}\mathcal{R}_u\right)Q_1$ in (60) is a stable matrix. Matrix stability means that all its eigenvalues should lie inside the unit circle. From the established statement on [21], all the entries of Q_1 and Q_2 are real non-negative and all the rows of Q_1 and Q_2 add up to unity, we know that

$Q_2 (I_{NM} - \mathcal{M}\mathcal{R}_u) Q_1$ is stable if the matrix $(I_{NM} - \mathcal{M}\mathcal{R}_u)$. or

$$|\lambda_{max} \{I_{NM} - \mathcal{M}\mathcal{R}_u\}| < 1 \tag{65}$$

It is know simple and easy to understand confirm that condition (65) ensures the stability of $(I_{NM} - \mathcal{M}\mathcal{R}_u)$

■

B. Mean-Square Stability

It is not sufficient to ensure the stability of the weight-error vectors in the mean sense. We need to examine how small the error vectors become. Doing so, we perform a mean-square error analysis. The purpose of the analysis is to evaluate how the variances $\mathbb{E}\left[\|\tilde{\boldsymbol{w}}_{k,i}\|^2\right]$ evolve with time and what their steady-state values are, for each node k. Here, we use the notation $\|x\|^2_\Sigma$ to denote the weighted square quantity $x^*\Sigma x$, for any column vector x and matrix Σ.

From recursion (59), we introduce

$$\boldsymbol{\mathcal{B}}_i \triangleq \boldsymbol{\mathcal{A}}_{2,i} (I_{NM} - \mathcal{M}\boldsymbol{\mathcal{R}}_{u,i}) \boldsymbol{\mathcal{A}}_{1,i-1} \tag{66}$$

$$\boldsymbol{\mathcal{H}}_i \triangleq \boldsymbol{\mathcal{A}}_{2,i} (I_{NM} - \mathcal{M}\boldsymbol{\mathcal{R}}_{u,i}) \tag{67}$$

So, we can rewrite (59) as

$$\tilde{\boldsymbol{w}}_i = \boldsymbol{\mathcal{B}}_i \tilde{\boldsymbol{w}}_{i-1} - \boldsymbol{\mathcal{H}}_i \boldsymbol{v}^{(w)}_{i-1} - \boldsymbol{\mathcal{A}}_{2,i}\mathcal{M}\boldsymbol{s}_i - \boldsymbol{v}^{(\psi)}_i \tag{68}$$

Taking the squared weighted Euclidean norm of both sides of (68) and apply the expectation operator together with using *Remark* and *Assumptions* yield the following weighted variance relation:

$$\begin{aligned}\mathbb{E}\left[\|\tilde{\boldsymbol{w}}_i\|^2_\Sigma\right] &= \mathbb{E}\left[\tilde{\boldsymbol{w}}^*_{i-1}\boldsymbol{\mathcal{B}}^*_i \Sigma \boldsymbol{\mathcal{B}}_i \tilde{\boldsymbol{w}}_{i-1}\right] + \mathbb{E}\left[\boldsymbol{s}^*_i \mathcal{M}\boldsymbol{\mathcal{A}}^T_{2,i}\Sigma\boldsymbol{\mathcal{A}}_{2,i}\mathcal{M}\boldsymbol{s}_i\right] \\ &+ \mathbb{E}\left[\boldsymbol{v}^{*(w)}_{i-1}\boldsymbol{\mathcal{H}}^*_i\Sigma\boldsymbol{\mathcal{H}}_i\boldsymbol{v}^{(w)}_{i-1}\right] + \mathbb{E}\left[\boldsymbol{v}^{*(\psi)}_i\Sigma\boldsymbol{v}^{(\psi)}_i\right]\end{aligned} \tag{69}$$

Let us evaluate each of the expectations on the right-hand side. The first expectation is given by

$$\begin{aligned}\mathbb{E}\left[\tilde{\boldsymbol{w}}^*_{i-1}\boldsymbol{\mathcal{B}}^*_i\Sigma\boldsymbol{\mathcal{B}}_i\tilde{\boldsymbol{w}}_{i-1}\right] &= \mathbb{E}\left[\mathbb{E}\left[\tilde{\boldsymbol{w}}^*_{i-1}\boldsymbol{\mathcal{B}}^*_i\Sigma\boldsymbol{\mathcal{B}}_i\tilde{\boldsymbol{w}}_{i-1}|\tilde{\boldsymbol{w}}_{i-1}\right]\right] \\ &= \mathbb{E}\left[\tilde{\boldsymbol{w}}^*_{i-1}\left(\mathbb{E}\left[\boldsymbol{\mathcal{B}}^*_i\Sigma\boldsymbol{\mathcal{B}}_i\right]\right)\right] \\ &= \mathbb{E}\left[\tilde{\boldsymbol{w}}^*_{i-1}\Sigma'\tilde{\boldsymbol{w}}_{i-1}\right] \\ &= \mathbb{E}\left[\|\tilde{\boldsymbol{w}}_i\|^2_{\Sigma'}\right]\end{aligned} \tag{70}$$

where we introduce the nonnegative-definite weighting matrix

$$\begin{aligned}\Sigma' &\triangleq \mathbb{E}\left[\boldsymbol{\mathcal{B}}^*_i\Sigma\boldsymbol{\mathcal{B}}_i\right] \\ &= \mathbb{E}\big[\boldsymbol{\mathcal{A}}^T_{1,i-1}(I_{NM} - \mathcal{M}\boldsymbol{\mathcal{R}}_{u,i})^*\boldsymbol{\mathcal{A}}^T_{2,i}\Sigma \\ &\quad \boldsymbol{\mathcal{A}}_{2,i}(I_{NM} - \mathcal{M}\boldsymbol{\mathcal{R}}_{u,i})\boldsymbol{\mathcal{A}}_{1,i-1}\big]\end{aligned} \tag{71}$$

It is convenient to introduce the alternative notation $\|x\|^2_\sigma$ to refer to the weighted square quantity $\|x\|^2_\Sigma$, where $\sigma = \mathrm{vec}\{\Sigma\}$. We shall use these two notations interchangeably.

Using the following equalities for arbitrary matrices $\{U, W, \Sigma, Z\}$ of compatible dimensions:

$$(U \otimes W)(\Sigma \otimes Z) = U\Sigma \otimes WZ \tag{72}$$

$$\mathrm{vec}\{U\Sigma W\} = \left(W^T \otimes U\right)\mathrm{vec}\{\Sigma\} \tag{73}$$

$$\mathrm{tr}(\Sigma W) = \left[\mathrm{vec}\left\{W^T\right\}\right]^T \mathrm{vec}\{\Sigma\} \tag{74}$$

we have

$$\begin{aligned}\sigma' &= \mathbb{E}\left[\boldsymbol{\mathcal{A}}^T_{1,i-1}\otimes\boldsymbol{\mathcal{A}}^T_{1,i-1}\right]\Big[(I_{NM} - \mathcal{M}\boldsymbol{\mathcal{R}}_{u,i})^* \otimes \\ &\quad (I_{NM} - \mathcal{M}\boldsymbol{\mathcal{R}}_{u,i})^T\Big]\mathbb{E}\left(\boldsymbol{\mathcal{A}}^T_{2,i}\otimes\boldsymbol{\mathcal{A}}^T_{2,i}\right)\mathrm{vec}\{\Sigma\} \\ &= \mathcal{D}_1\left[(I_{NM} - \mathcal{M}\boldsymbol{\mathcal{R}}_{u,i})^* \otimes (I_{NM} - \mathcal{M}\boldsymbol{\mathcal{R}}_{u,i})^T\right]\mathcal{D}_2\sigma\end{aligned} \tag{75}$$

where

$$\sigma = \mathrm{vec}\{\Sigma\} \tag{76}$$

$$\sigma' = \mathrm{vec}\{\Sigma'\} \tag{77}$$

$$\mathcal{D}_1 = \mathbb{E}\left[\boldsymbol{\mathcal{A}}^T_{1,i-1}\otimes\boldsymbol{\mathcal{A}}^T_{1,i-1}\right] \tag{78}$$

$$\mathcal{D}_2 = \mathbb{E}\left[\boldsymbol{\mathcal{A}}^T_{2,i}\otimes\boldsymbol{\mathcal{A}}^T_{2,i}\right] \tag{79}$$

In Appendix B, we calculate $\mathcal{D}_1$ and $\mathcal{D}_2$ for both stochastic and sequential partial-diffusion schemes. That is,

$$\sigma' = \mathcal{F}\sigma \tag{80}$$

where we are introducing the coefficient matrix of size $(NM)^2 \times (NM)^2$:

$$\mathcal{F} = \mathcal{D}_1\left[\mathbb{E}(I_{NM} - \mathcal{M}\boldsymbol{\mathcal{R}}_{u,i})^* \otimes (I_{NM} - \mathcal{M}\boldsymbol{\mathcal{R}}_{u,i})^T\right]\mathcal{D}_2 \tag{81}$$

Second term on RHS of (69)

$$\mathbb{E}\left(\boldsymbol{s}^*_i\mathcal{M}\boldsymbol{\mathcal{A}}^T_{2,i}\Sigma\boldsymbol{\mathcal{A}}_{2,i}\mathcal{M}\boldsymbol{s}_i\right) = \mathrm{vec}^T\{\mathcal{G}\}\mathcal{D}_2\sigma \tag{82}$$

where

$$\mathcal{G} = \mathcal{M}\mathbb{E}\left[\boldsymbol{s}_i\boldsymbol{s}^*_i\right]\mathcal{M} \tag{83}$$

By considering *Assumptions*, $\mathcal{G}$ is evaluated as

$$\mathcal{G} = \left\{\mu_1\sigma^2_{v,1}R_{u,1}, \ldots, \mu_N\sigma^2_{v,N}R_{u,N}\right\} \tag{84}$$

Third term on RHS of (69)

$$\begin{aligned}&\mathbb{E}\left[\boldsymbol{v}^{*(w)}_{i-1}\boldsymbol{\mathcal{H}}^*_i\Sigma\boldsymbol{\mathcal{H}}_i\boldsymbol{v}^{(w)}_{i-1}\right] \\ &= \mathbb{E}\Big[\boldsymbol{v}^{*(w)}_{i-1}(I_{NM} - \boldsymbol{\mathcal{R}}_{u,i}\mathcal{M})\boldsymbol{\mathcal{A}}^T_{2,i}\times \\ &\qquad \Sigma\boldsymbol{\mathcal{A}}_{2,i}(I_{NM} - \mathcal{M}\boldsymbol{\mathcal{R}}_{u,i})\boldsymbol{v}^{(w)}_{i-1}\Big] \\ &= \mathbb{E}\Big[\mathrm{tr}\Big(\Sigma\boldsymbol{\mathcal{A}}_{2,i}(I_{NM} - \mathcal{M}\boldsymbol{\mathcal{R}}_{u,i}) \\ &\qquad \boldsymbol{v}^{(w)}_{i-1}\boldsymbol{v}^{*(w)}_{i-1}(I_{NM} - \boldsymbol{\mathcal{R}}_{u,i}\mathcal{M})\boldsymbol{\mathcal{A}}^T_{2,i}\Big)\Big] \\ &= \mathbb{E}\Big[\mathrm{vec}^T\Big\{\boldsymbol{\mathcal{A}}_{2,i}(I_{NM} - \mathcal{M}\boldsymbol{\mathcal{R}}_{u,i})\times \\ &\qquad \boldsymbol{v}^{(w)}_{i-1}\boldsymbol{v}^{*(w)}_{i-1}(I_{NM} - \boldsymbol{\mathcal{R}}_{u,i}\mathcal{M})\boldsymbol{\mathcal{A}}^T_{2,i}\Big\}\Big]^T\mathrm{vec}\{\Sigma\} \\ &= \mathrm{vec}^T\Big\{(I_{NM} - \mathcal{M}\mathcal{R}_u)R^{(w)}_v\times \\ &\qquad (I_{NM} - \mathcal{M}\mathcal{R}_u)\Big\}\mathbb{E}\left[\boldsymbol{\mathcal{A}}^T_{2,i}\otimes\boldsymbol{\mathcal{A}}^T_{2,i}\right]\sigma \\ &= \mathrm{vec}^T\left\{\mathcal{U}R^{(w)}_v\mathcal{U}^*\right\}\mathcal{D}_2\sigma\end{aligned} \tag{85}$$

where

$$\mathcal{U} \triangleq \mathbb{E}\left[I_{NM} - \mathcal{M}\boldsymbol{\mathcal{R}}_{u,i}\right] = \left(I_{NM} - \mathcal{M}\mathcal{R}_u\right) \tag{86}$$

Last term on RHS of (69):

$$\begin{aligned}\mathbb{E}\left[\boldsymbol{v}_i^{*(\psi)}\Sigma\boldsymbol{v}_i^{(\psi)}\right] &= \mathbb{E}\left[\mathrm{tr}\left(\Sigma\boldsymbol{v}_i^{(\psi)}\boldsymbol{v}_i^{*(\psi)}\right)\right] \\ &= \mathrm{vec}^T\left\{R_v^{(\psi)}\right\}\sigma\end{aligned} \tag{87}$$

The variance relation becomes

$$\begin{aligned}\mathbb{E}\left[\|\tilde{\boldsymbol{w}}_i\|_\sigma^2\right] &= \mathbb{E}\left[\|\tilde{\boldsymbol{w}}_i\|_{\mathcal{F}\sigma}^2\right] \\ &+ \Bigg(\mathrm{vec}^T\{\mathcal{G}\}\mathcal{D}_2 + \mathrm{vec}^T\{\mathcal{U}R_v^{(w)}\mathcal{U}^*\}\mathcal{D}_2 \\ &+ \mathrm{vec}^T\{R_v^{(\psi)}\}\Bigg)\sigma\end{aligned} \tag{88}$$

Theorem 2 (Mean-Square Stability). Consider the same setting of *Theorem* 1. Assume sufficiently small step-sized to justify ignoring terms that depend on higher power of the step-sizes. The perturbed adaptive partial diffusion algorithm (23)-(25) is mean-square stable if, and only if, the matrix $\mathcal{F}$ defined by (81), or its approximate defines further (89), is stable. This condition is satisfied small for step-sizes $\{\mu_k\}$ as is (64).

Proof: A resonable approximate expression for $\mathcal{F}$ for sufficiently small step-sizes is

$$\mathcal{F} \approx \mathcal{D}_1\left[\left(I_{NM} - \mathcal{M}\mathcal{R}_u\right)^T \otimes \left(I_{NM} - \mathcal{M}\mathcal{R}_u\right)^T\right]\mathcal{D}_2 \tag{89}$$

Recall that, in the Kronecker product case $(C = B \otimes A)$, eigenvalues are the outer product of the eigenvalues of the two matrices. Therefore, using expression (89), we have that $\zeta\left(\mathcal{F}\right) = \left[\zeta\left(I_{NM} - \mathcal{M}\mathcal{R}_u\right)\right]$, where $\zeta\left(A\right)$ is the spectral radius A. It follows that $\mathcal{F}$ is stable if, and only if, $\left(I_{NM} - \mathcal{M}\mathcal{R}_u\right)$ is stable. we already mentioned in (1) that (65) ensures the stability of $\left(I_{NM} - \mathcal{M}\mathcal{R}_u\right)$.

■

Corollary (Steady-State Variance Relation). Consider the same setting of Theorem 2. The weight-error vector, $\tilde{\boldsymbol{w}}_i$, of the (PDLMS) (23)-(25) satisfies the following equation in steady-state:

$$\begin{aligned}\mathbb{E}\left[\|\tilde{\boldsymbol{w}}_i\|_{\left(I_{N^2M^2} - \mathcal{F}\right)\sigma}\right] &= \\ \Bigg(\mathrm{vec}^T\{\mathcal{G}\}\mathcal{D}_2 &+ \mathrm{vec}^T\{\mathcal{U}R_v^{(w)}\mathcal{U}^*\}\mathcal{D}_2 \\ &+ \mathrm{vec}^T\{R_v^{(\psi)}\}\Bigg)\sigma\end{aligned} \tag{90}$$

for any Hermitian nonnegative-definite matrix Σ which follows (76).

C. Mean-Square Performance

Expression (90) prove a very useful relation; it allows us to evaluate the network MSD through appropriate selection of the weighting matrix Σ. The network MSD is defined as the average value:

$$\mathrm{MSD}^{\mathrm{network}} \triangleq \lim_{i\to\infty}\frac{1}{N}\sum_{k=1}^{N}\mathbb{E}\left[\|\tilde{\boldsymbol{w}}_{k,i}\|^2\right] \tag{91}$$

which amounts to averaging the MSDs of the individual nodes. Therefore,

$$\mathrm{MSD}^{\mathrm{network}} = \lim_{i\to\infty}\frac{1}{N}\mathbb{E}\left[\|\tilde{\boldsymbol{w}}_i\|^2\right] = \lim_{i\to\infty}\mathbb{E}\left[\|\tilde{\boldsymbol{w}}_i\|_{1/N}^2\right] \tag{92}$$

This means that in order to recover the network MSD from (90), we should select the weighting vector σ such that

$$\left(I_{N^2M^2} - \mathcal{F}\right)\sigma = \frac{1}{N}\mathrm{vec}\left\{I_{NM}\right\} \tag{93}$$

Solving for σ and substituting back into (90) we arrive at the following expression for the network MSD

$$\begin{aligned}\mathrm{MSD}_{\mathrm{noisy}}^{\mathrm{network}} &= \frac{1}{N}\Bigg(\mathrm{vec}^T\{\mathcal{G}\}\mathcal{D}_2 + \mathrm{vec}^T\{\mathcal{U}R_v^{(w)}\mathcal{U}^*\}\mathcal{D}_2 \\ &+ \mathrm{vec}^T\{R_v^{(\psi)}\}\Bigg) \times \left(I_{N^2M^2} - \mathcal{F}\right)^{-1}\mathrm{vec}\left\{I_{NM}\right\}\end{aligned} \tag{94}$$

When links are ideal, the last two terms of (90) do not arise. So, we can conclude that the network MSD deteriorates as follows:

$$\begin{aligned}\mathrm{MSD}_{\mathrm{noisy}}^{\mathrm{network}} &= \mathrm{MSD}_{\mathrm{ideal}}^{\mathrm{network}} + \\ &\frac{1}{N}\left(\mathrm{vec}^T\{\mathcal{U}R_v^{(w)}\mathcal{U}^*\}\mathcal{D}_2 + \mathrm{vec}^T\{R_v^{(\psi)}\}\right) \times \\ &\left(I_{N^2M^2} - \mathcal{F}\right)^{-1}\mathrm{vec}\left\{I_{NM}\right\}\end{aligned} \tag{95}$$

IV. Detailed Discussion on the Network MSD

So far we have mentioned based on Theorems 1 on 2 that the partial diffusion LMS strategy does not diverge due to noisy links. But, it is the main factor on performance degradation of steady-state network MSD. Moreover, focus on (95) and compare it with that stated at [21], there exist an additional term, denoted as channel noise term, that plays a crucial rule on the performance degradation of network MSD performance. It is abundantly clear that this term has been arose from channel noise condition. Here we concentrate on (95) to explicitly highlight characterization of how convergence and performance of PDLMS affects by the presence of noisy channels. We analyze steady-state network MSD under the following assumptions:

Assumptions.

(v) Nodes run ATC PDLMS at each iteration, i.e., $C_1 = I_N$. For the sake of simplicity of notation we consider, $C_2 = C$ and $\mathcal{D}_2 = \mathcal{D}$.

(vi) The step size, noise variance, input covariance matrix and channel noise covariance matrix all are the same in the network, i.e., $\mu_k = \mu$, $\sigma_{v,k}^2 = \sigma_v^2$, $R_{u,k} = R_u$, and $R_{v,k}^{(\psi)} = R_v^{(\psi)}$, $\forall k \in 1, \ldots, N$.

(vii) During any M consecutive iterations, the intermediate estimate vector does not differ considerably at each node k [21].

Considering *Assumptions* (v), (vii), and using the results of analysis mentioned in [21], we have

$$\mathcal{D} = \left(1 - \rho\right)I_{N^2M^2} + \rho\,\mathcal{C}\otimes\mathcal{C} \tag{96}$$

where, $\mathcal{C} = C \otimes I_M$.

To specify the network performance in steady-state, we consider the global MSD describe at (95) and denote it by η_L. We have

$$\eta_L = \frac{1}{N}\left(\mathrm{vec}^T\{\mathcal{G}\}\mathcal{D} + \mathrm{vec}^T\{R_v^{(\psi)}\}\right)\left(\sum_{n=0}^{\infty}\mathcal{F}^n\right)\mathrm{vec}\{I_{NM}\}$$
$$= \frac{1}{N}\sum_{n=0}^{\infty}\left(\mathrm{vec}^T\{\mathcal{G}\}\mathcal{D} + \mathrm{vec}^T\{R_v^{(\psi)}\}\right) \times \left[\left(\mathcal{U}^T \otimes \mathcal{U}^T\right)\mathcal{D}\right]^n \mathrm{vec}\{I_{NM}\} \quad (97)$$

Substituting (96) into (97) results in

$$\eta_L = \frac{1}{N}\sum_{n=0}^{\infty}(1-\rho)^{n+1}\mathrm{vec}^T\{\mathcal{G}\}\left[(\mathcal{U})^n \otimes (\mathcal{U})^n\right]^T \times \mathrm{vec}\{I_{NM}\}$$
$$+ \frac{1}{N}\sum_{n=0}^{\infty}\rho^{n+1}\mathrm{vec}^T\{\mathcal{G}\}\left[\mathcal{C}\left(\mathcal{U}^T\mathcal{C}\right)^n \otimes \mathcal{C}\left(\mathcal{U}^T\mathcal{C}\right)^n\right] \times \mathrm{vec}\{I_{NM}\}$$
$$+ \frac{1}{N}\sum_{n=0}^{\infty}(1-\rho)^n\mathrm{vec}^T\left\{R_v^{(\psi)}\right\}\left[(\mathcal{U})^n \otimes (\mathcal{U})^n\right]^T \times \mathrm{vec}\{I_{NM}\}$$
$$+ \frac{1}{N}\sum_{n=0}^{\infty}\rho^n\mathrm{vec}^T\left\{R_v^{(\psi)}\right\}\left[\mathcal{C}\left(\mathcal{U}^T\mathcal{C}\right)^n \otimes \mathcal{C}\left(\mathcal{U}^T\mathcal{C}\right)^n\right] \times \mathrm{vec}\{I_{NM}\} \quad (98)$$

Utilizing $\mathrm{vec}\{.\}$ property, (98) can be rewrite as

$$\eta_L = \frac{1}{N}\sum_{n=0}^{\infty}(1-\rho)^{n+1}\mathrm{vec}^T\left\{\left[(\mathcal{U})^n\mathcal{G}^T(\mathcal{U})^n\right]^T\right\} \times \mathrm{vec}\{I_{NM}\}$$
$$+ \frac{1}{N}\sum_{n=0}^{\infty}\rho^{n+1}\mathrm{vec}^T\left\{\left[\mathcal{C}\left(\mathcal{U}^T\mathcal{C}\right)^n\mathcal{G}\mathcal{C}\left(\mathcal{U}^T\mathcal{C}\right)^n\right]\right\} \times \mathrm{vec}\{I_{NM}\}$$
$$+ \frac{1}{N}\sum_{n=0}^{\infty}(1-\rho)^n\mathrm{vec}^T\left\{\left[(\mathcal{U})^n\left(R_v^{(\psi)}\right)^T(\mathcal{U})^n\right]^T\right\} \times \mathrm{vec}\{I_{NM}\}$$
$$+ \frac{1}{N}\sum_{n=0}^{\infty}\rho^n\mathrm{vec}^T\left\{\left[\mathcal{C}\left(\mathcal{U}^T\mathcal{C}\right)^n R_v^{(\psi)}\mathcal{C}\left(\mathcal{U}^T\mathcal{C}\right)^n\right]\right\} \times \mathrm{vec}\{I_{NM}\} \quad (99)$$

Using trace properties, we have

$$\eta_L = \frac{1}{N}\sum_{n=0}^{\infty}(1-\rho)^{n+1}\mathrm{tr}\left\{\left[(\mathcal{U})^n\mathcal{G}^T(\mathcal{U})^n\right]^T\right\}$$
$$+ \frac{1}{N}\sum_{n=0}^{\infty}\rho^{n+1}\mathrm{tr}\left\{\left[\mathcal{C}\left(\mathcal{U}^T\mathcal{C}\right)^n\mathcal{G}\mathcal{C}\left(\mathcal{U}^T\mathcal{C}\right)^n\right]\right\}$$
$$+ \frac{1}{N}\sum_{n=0}^{\infty}(1-\rho)^n\mathrm{tr}\left\{\left[(\mathcal{U})^n\left(R_v^{(\psi)}\right)^T(\mathcal{U})^n\right]^T\right\}$$
$$+ \frac{1}{N}\sum_{n=0}^{\infty}\rho^n\mathrm{tr}\left\{\left[\mathcal{C}\left(\mathcal{U}^T\mathcal{C}\right)^n R_v^{(\psi)}\mathcal{C}\left(\mathcal{U}^T\mathcal{C}\right)^n\right]\right\} \quad (100)$$

In view of *Assumption* (v), we have

$$\mathcal{M} = \mu I_N \otimes I_M,$$
$$\mathcal{U} = I_N \otimes (I_M - \mu R_u),$$
$$\mathcal{G} = \mu^2\sigma^2 I_M \otimes R_u,$$

and

$$R_v^{(\psi)} = R_v$$

Calculating the trace terms at (100), we get

$$\mathrm{tr}\left\{\left[(\mathcal{U})^n\mathcal{G}^T(\mathcal{U})^n\right]^T\right\} = N\mu^2\sigma^2 \times \mathrm{tr}\left\{\left[(I_N - \mu R_u)^n R_u (I_N - \mu R_u)^n\right]^T\right\},$$

$$\mathrm{tr}\left\{\left[\mathcal{C}\left(\mathcal{U}^T\mathcal{C}\right)^n\mathcal{G}\mathcal{C}\left(\mathcal{U}^T\mathcal{C}\right)^n\right]\right\} = \mu^2\sigma_v^2\mathrm{tr}\left\{\left(C^T\right)^{n+1}C^{n+1}\right\} \times \mathrm{tr}\left\{(I_M - \mu R_u)^n R_u (I_M - \mu R_u)^n\right\},$$

$$\mathrm{tr}\left\{\left[(\mathcal{U})^n\left(R_v^{(\psi)}\right)^T(\mathcal{U})^n\right]^T\right\} = N \times \mathrm{tr}\left\{\left[(I_N - \mu R_u)^n (R_v)^T (I_N - \mu R_u)^n\right]^T\right\},$$

and

$$\mathrm{tr}\left\{\left[\mathcal{C}\left(\mathcal{U}^T\mathcal{C}\right)^n R_v^{(\psi)}\mathcal{C}\left(\mathcal{U}^T\mathcal{C}\right)^n\right]\right\} = \mathrm{tr}\left\{\left(C^T\right)^{n+1}C^{n+1}\right\} \times \mathrm{tr}\left\{(I_M - \mu R_u)^n R_v (I_M - \mu R_u)^n\right\}$$

where

$$\mathrm{tr}\left\{\left(C^T\right)^{n+1}C^{n+1}\right\} = \sum_{k=1}^{N}\|c_{n+1,k}\|^2, n \geq 0$$

In equation above, $c_{n+1,k}$ is the k-th column of C^{n+1}. Considering a connected network holds that,

$$\|c_{n+1,k}\|^2 < 1, n \geq 0, \forall k.$$

This fact yields

$$\mathrm{tr}\left\{\left[\mathcal{C}\left(\mathcal{U}^T\mathcal{C}\right)^n\mathcal{G}\mathcal{C}\left(\mathcal{U}^T\mathcal{C}\right)^n\right]\right\} < \mathrm{tr}\left\{\left[(\mathcal{U})^n\mathcal{G}^T(\mathcal{U})^n\right]^T\right\}$$
$$< \mathrm{tr}\left\{\left[\mathcal{C}\left(\mathcal{U}^T\mathcal{C}\right)^n\mathcal{G}\mathcal{C}\left(\mathcal{U}^T\mathcal{C}\right)^n\right]\right\} + \mathrm{tr}\left\{\left[\mathcal{C}\left(\mathcal{U}^T\mathcal{C}\right)^n R_v^{(\psi)}\mathcal{C}\left(\mathcal{U}^T\mathcal{C}\right)^n\right]\right\} \quad (101)$$

Moreover, for the case of non-cooperative, i.e. $L = 0$ and $\rho = 0$,, full diffusion case with ideal links, i.e. $L = M$, $\rho = 1$, and $R_v = O_M$, full diffusion case with noisy links, i.e. $L = M$, $\rho = 1$ and $R_v \neq O_M$, where O_M is and $M \times M$ zero matrix, we have

$$\eta_0 = \frac{1}{N}\sum_{n=0}^{\infty}\mathrm{tr}\left\{\left[(\mathcal{U})^n\mathcal{G}^T(\mathcal{U})^n\right]^T\right\}, \quad (102)$$

$$\eta_M = \frac{1}{N}\sum_{n=0}^{\infty}\mathrm{tr}\left\{\left[\mathcal{C}\left(\mathcal{U}^T\mathcal{C}\right)^n\mathcal{G}\mathcal{C}\left(\mathcal{U}^T\mathcal{C}\right)^n\right]\right\} \quad (103)$$

and

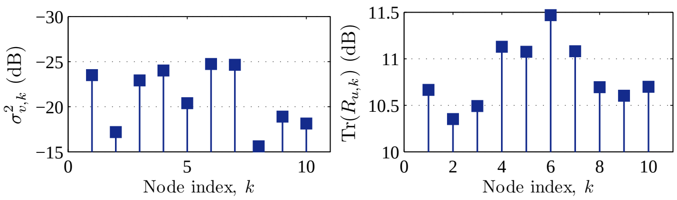


Fig. 1. Variance of the noise (top) and Covariance matrix trace of the input signal (bottom) at each node.

$$\eta_M = \frac{1}{N}\sum_{n=0}^{\infty}\Bigg(\mathrm{tr}\Big\{\Big[\mathcal{C}\left(\mathcal{U}^T\mathcal{C}\right)^n \mathcal{G}\mathcal{C}\left(\mathcal{U}^T\mathcal{C}\right)^n\Big]\Big\} + \mathrm{tr}\Big\{\Big[\mathcal{C}\left(\mathcal{U}^T\mathcal{C}\right)^n R_v^{(\psi)}\mathcal{C}\left(\mathcal{U}^T\mathcal{C}\right)^n\Big]\Big\}\Bigg) \quad (104)$$

respectively. Accordingly, it follows from (97)-(104)

$$\eta_M < \eta_{M-1} < \ldots < \eta_L < \ldots < \eta_1 < \eta_0 < \eta_1^{noisy} < \ldots < \eta_L^{noisy} < \ldots < \eta_{M-1}^{noisy} < \eta_M^{noisy}$$

This indicates that under ideal channels the more entries are transmitted at each iteration, the better the steady-state network MSD performance occurred. In other words, partial diffusion settles a communication performance trade-off. However, considering noisy channels the more entries are communicated at each iteration, the worse the steady-state network MSD performance occurred. This means that partial diffusion seriously upset the balance of mentioned trade-off.

V. Simulation Results

In order to illustrate the PLDMS performance under noisy information exchange, we consider an adaptive network with a random topology and $N = 10$ where each node is, on average, connected to two other nodes. The measurements were generated according to model (1), and regressors, $\boldsymbol{u}_{k,i}$, were chosen Gaussian i.i.d with randomly generated different diagonal covariance matrices, $R_{u,k}$. The additive noises at the nodes are zero mean Gaussian with variances $\sigma^2_{v,k}$ and independent of the regression data. The traces of the covariance matrix regressors and the noise variances at all nodes, $\mathrm{tr}\left(R_{u,k}\right)$ and $\sigma^2_{v,k}$, are shown in Fig. 1. We also use white Gaussian link noise signals such that $R^{(w)}_{w,lk} = \sigma^2_{v,lk}I_M$ and $R^{(\psi)}_{v,lk} = \sigma^2_{\psi,lk}I_M$. All link noise variances $\{\sigma^2_{w,lk}, \sigma^2_{\psi,lk}\}$ are randomly generated and illustrated in Fig. 2.

We assign the link number by the following procedure. We denote the link from node l to node k as $\ell_{l,k}$, where $l \neq k$. Then, we sort the links $\{\ell_{l,k}, l \in \mathcal{N}_k \backslash \{k\}\}$ in an ascending order of l in the list $\mathcal{L}_k$ (which is a set with ordered elements) for each node. We concatenate $\{\mathcal{L}_k\}$ in an ascending order of k to obtain the overall list $\mathcal{L} = \{\mathcal{L}_1, \mathcal{L}_2, \ldots, \mathcal{L}_N\}$. Eventually, the mth link in the network is given by the mth element in the list $\mathcal{L}$. We adopt different step-sizes μ_k, randomly generated, but within the limits in (65) for all agents. It is noteworthy that we adopt the network MSD learning curves of all figures by averaging over 50 experiments and the unknown parameter w^o of length $M = 8$ is randomly generated.

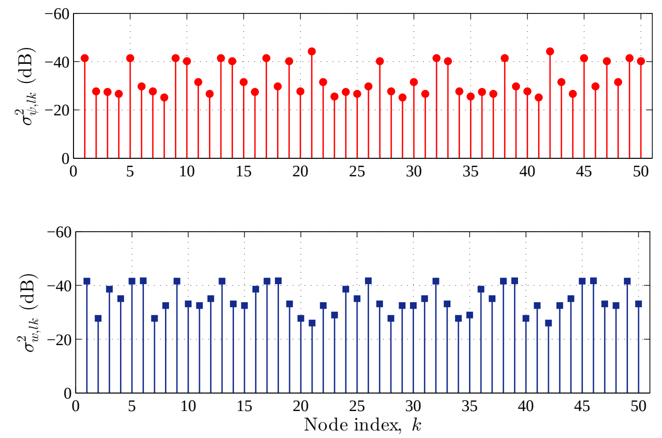


Fig. 2. The variance profiles for various sources of link noises in dB, including $\left\{\sigma^2_{\psi,lk}, \sigma^2_{w,lk}\right\}$, $\sigma^2_{\psi,lk}$ (top) and $\sigma^2_{w,lk}$ (bottom).

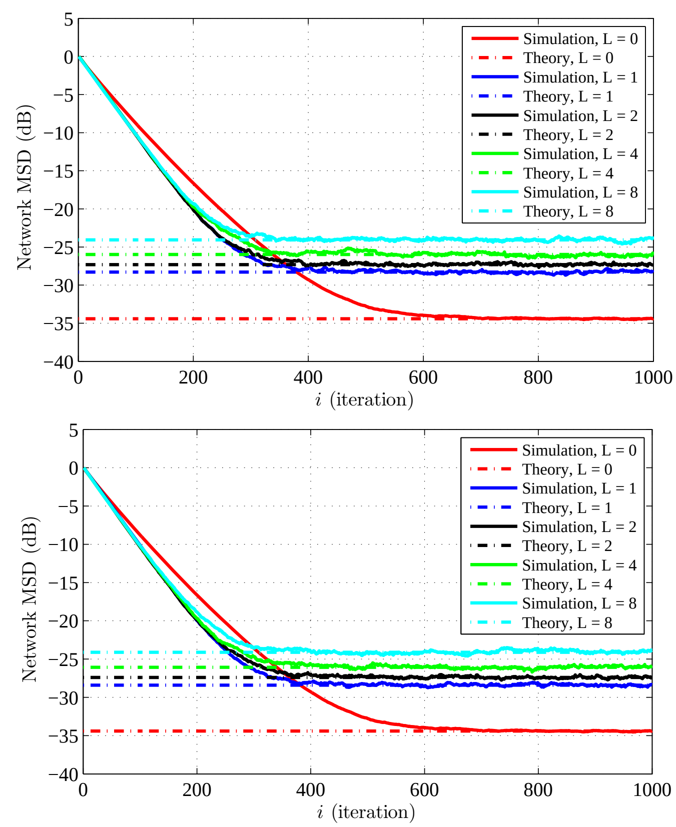


Fig. 3. Simulated network MSD curves and theoretical results (95) for partial diffusion ATC algorithms using sequential (top) and stochastic (bottom) with different number of entries communicated under noisy links.

In the first simulation, we evaluate the theoretical derivations. To this end we consider the experimental network MSD learning curves (ATC strategy) of PDLMS algorithm and theoretical results using both sequential and stochastic partial diffusion schemes under noisy links for different numbers of entries, L. We use uniform weights for $\{c_{1,lk}, c_{2,lk}\}$ at the combination phase at this stage. The plots are given in Fig. 3 where we can see that there is a good match between our theoretical derivations with simulation results. Similar plots

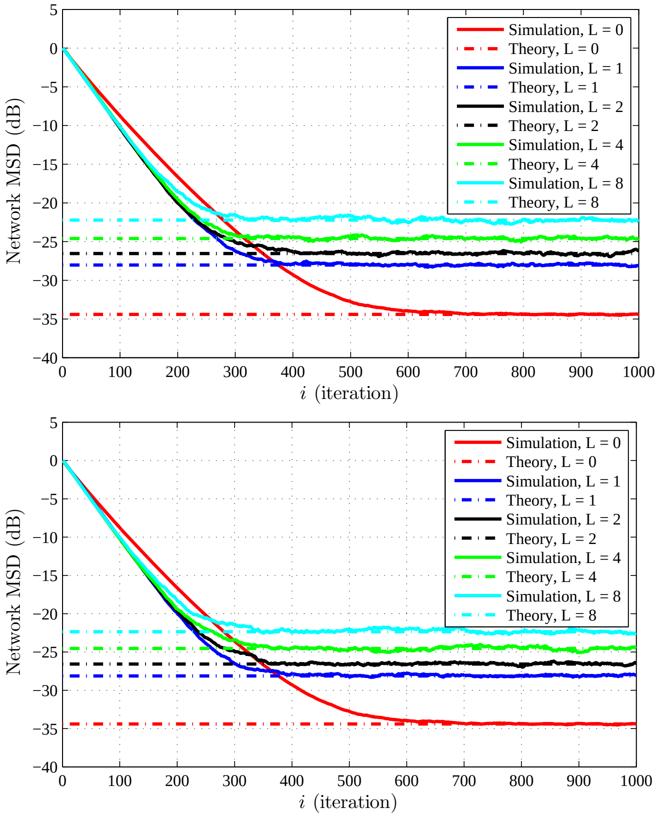


Fig. 4. Simulated network MSD curves and theoretical results (95) for partial diffusion CTA algorithms using sequential (top) and stochastic (bottom) with different number of entries communicated under noisy links.

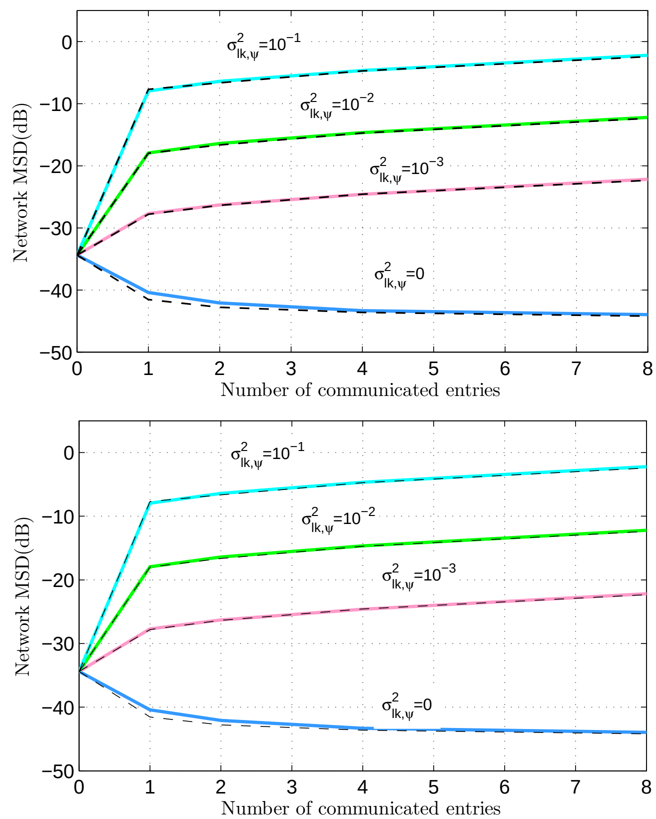


Fig. 5. Theoretical and experimental steady-state network MSDs versus the number of entries communicated at each iteration for different values of $\sigma^2_{\psi,lk}$ using sequential (top) and stochastic (bottom). Note that solid line and dashed line represent the theoretical and experimental results respectively.

for CTA strategy are given in Fig. 4.

To further examine our theoretical findings, both theoretical and experimental steady-state network MSD of the ATC PDLMS algorithm as a function of communicated entries, L, for different values of $\{\sigma^2_{\psi,lk}\}$, for stochastic and sequential schemes are plotted in Fig. 5. This figure not only supports our analysis, but it also reveals that when channel between agents is assumed ideal ($\sigma^2_{\psi,lk} = 0$ in the figure), an increase in communicated entries results in the network performance. It must be noted that this in not the case when links among the nodes are noisy and the performance of network is deeply affected by variance of channel noise. This particular behaviour of the PDLMS algorithm in the presence of noisy links is better understood from Fig. 6 where the steady-state MSDs of all node k for different values of L and different link conditions are plotted in Fig. 6.

It is also notable that in the presence of noisy links, the PDLMS algorithm exhibits different behaviour as the step size changes. To show this behaviour, the steady-state MSD as a function of μ for different values of $\sigma^2_{\psi,lk}$ is shown in Fig. 7. As it is obvious from Fig. 7, for the case of ideal links $\sigma^2_{\psi,lk} = 0$ the MSD curve is a monotonically increasing function of μ [14], whereas, for noisy links, decreasing the step size increases the steady-state MSD value. Also, we can see from Fig. 7 that as $\sigma^2_{\psi,lk}$ increases, the effect of noisy links increases as expected. Finally, it must be noted that although the performance of PDLMS algorithm deteriorates in the presence of noisy links, it is still able to provide to deliver better performance in comparison with some similar methods, such as a consensus based algorithm. To show this, the MSD performance for different algorithms including non-cooperative, consensus, full diffusion and partial-diffusion (for $L = 2$ and $L = 4$) under noisy links is illustrated in Fig. 8. We can observe that the DLMS algorithm exhibits better performance than the consensus algorithm.

From the results above, we can make the following observations:

- The PDLMS algorithm delivers a trade-off between communications cost and estimation performance under ideal link. This statement is the main aim of [21].
- Unlike the statement above that labels a communication-performance trade-off to PDLMS algorithm, there is no direct relation between MSD performance and number of selected entries under noisy information exchange. In other words, the more entries are communicated at each iteration, the more perturbed weight estimates are interred in consultant phase that leads to worse steady-state network MSD.
- The sequential partial-diffusion schemes outperform the stochastic partial-diffusion for noisy and ideal links.
- The ATC PDLMS strategy outperforms the adaptive CTA PDLMS strategy for both noisy and ideal cases.

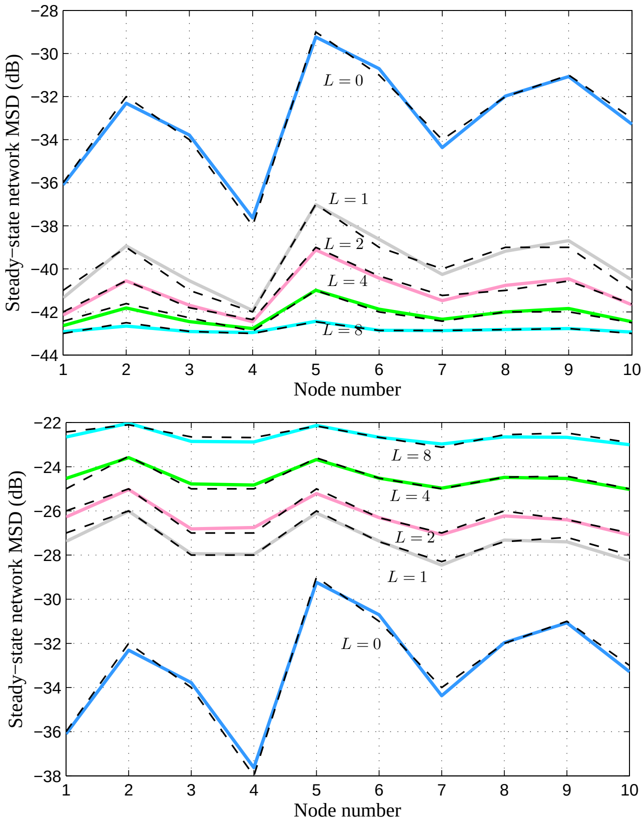


Fig. 6. Theoretical and experimental steady-state MSDs at each node for different numbers of entries communicated at each iteration under ideal links (top) and noisy links (bottom). Note that solid line and dashed line represent the theoretical and experimental results respectively.

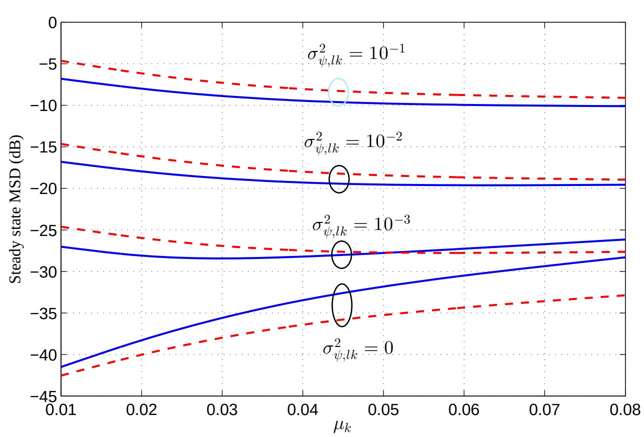


Fig. 7. The MSD of ATC-PDLMS as a function of μ for different values of $\sigma^2_{\psi,lk}$.

VI. Conclusion

In this work, we present a general form of PDLMS algorithms, formulate the ATC and CTA version of PDLMS under noisy links condition, and investigate the performance of partial-diffusion algorithms under several sources of noise during information exchange for both sequential and stochastic cases. We also illustrate that the PDLMS strategy can still stabilize the mean and mean-square convergence of the network with non-ideal links. We derived analytical expressions for network learning curve MSD. Furthermore, we established that there is not a direct relation between the MSD performance and the number of selected entries under imperfect information exchange. In other words, the more entries are communicated at each iteration, the more perturbed weight estimates are interred in consultant phase. The simulation results verify the theoretical findings and how well they match with theory.

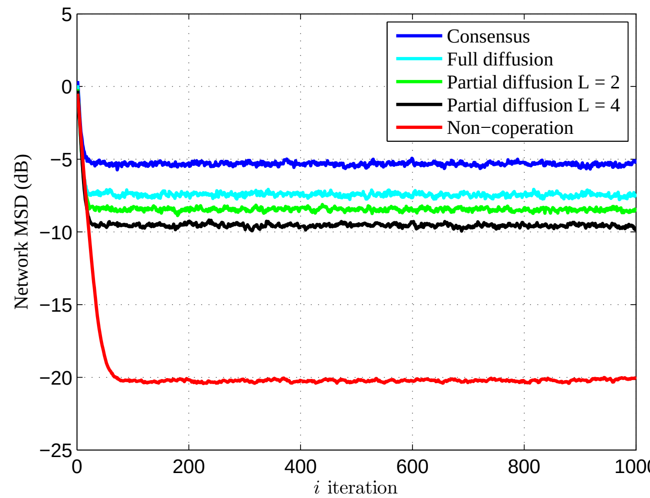


Fig. 8. Comparison of network MSDs for: non-cooperative, consensus, full diffusion, and partial diffusion strategies.

References

[1] H. Sayama, I. Pestov, J. Schmidt, B. J. Bush, C. Wong, J. Yamanoi, and T. Gross, “Modeling complex systems with adaptive networks,” *Computers & Mathematics with Applications*, vol. 65, no. 10, pp. 1645–1664, 2013.

[2] A. H. Sayed, “Diffusion adaptation over networks,” *Academic Press Library in Signal Processing*, vol. 3, pp. 323–454, 2013.

[3] A. H. Sayed, S.-Y. Tu, J. Chen, X. Zhao, and Z. J. Towfic, “Diffusion strategies for adaptation and learning over networks: an examination of distributed strategies and network behavior,” *Signal Processing Magazine, IEEE*, vol. 30, no. 3, pp. 155–171, 2013.

[4] R. Olfati-Saber, A. Fax, and R. M. Murray, “Consensus and cooperation in networked multi-agent systems,” *Proceedings of the IEEE*, vol. 95, no. 1, pp. 215–233, 2007.

[5] C. Lopes and A. H. Sayed, “Distributed processing over adaptive networks,” in *Proc. adaptive sensor array processing workshop*, 2006, pp. 1–5.

[6] C. G. Lopes and A. H. Sayed, “Diffusion least-mean squares over adaptive networks: Formulation and performance analysis,” *Signal Processing, IEEE Transactions on*, vol. 56, no. 7, pp. 3122–3136, 2008.

[7] F. S. Cattivelli and A. H. Sayed, “Diffusion LMS algorithms with information exchange,” in *2008 42nd Asilomar Conference on Signals, Systems and Computers*, 2008.

[8] ——, “Diffusion LMS strategies for distributed estimation,” *Signal Processing, IEEE Transactions on*, vol. 58, no. 3, pp. 1035–1048, 2010.

[9] ——, “Diffusion strategies for distributed Kalman filtering and smoothing,” *Automatic Control, IEEE Transactions on*, vol. 55, no. 9, pp. 2069–2084, 2010.

[10] F. S. Cattivelli, C. G. Lopes, and A. H. Sayed, “Diffusion recursive least-squares for distributed estimation over adaptive networks,” *Signal Processing, IEEE Transactions on*, vol. 56, no. 5, pp. 1865–1877, 2008.

[11] S. Xie and H. Li, “Distributed lms estimation over networks with quantised communications,” *International Journal of Control*, vol. 86, no. 3, pp. 478–492, 2013.

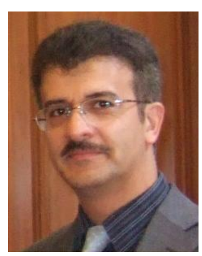
[12] R. Abdolee and B. Champagne, “Diffusion LMS algorithms for sensor networks over non-ideal inter-sensor wireless channels,” in *Distributed Computing in Sensor Systems and Workshops (DCOSS), 2011 International Conference on*. IEEE, 2011, pp. 1–6.

[13] A. Khalili, M. A. Tinati, A. Rastegarnia, and J. A. Chambers, "Transient analysis of diffusion least-mean squares adaptive networks with noisy channels," *International Journal of Adaptive Control and Signal Processing*, vol. 26, no. 2, pp. 171–180, 2012.
[14] A. Khalili, M. A. Tinati, A. Rastegarnia, and J. Chambers, "Steady-state analysis of diffusion LMS adaptive networks with noisy links," *Signal Processing, IEEE Transactions on*, vol. 60, no. 2, pp. 974–979, 2012.
[15] S.-Y. Tu and A. H. Sayed, "Adaptive networks with noisy links," in *Global Telecommunications Conference (GLOBECOM 2011), 2011 IEEE*. IEEE, 2011, pp. 1–5.
[16] A. Rastegarnia, W. M. Bazzi, A. Khalili, and J. A. Chambers, "Diffusion adaptive networks with imperfect communications: link failure and channel noise," *IET Signal Processing*, vol. 8, no. 1, pp. 59–66, Feb 2014.
[17] R. Abdolee and B. Champagne, "Diffusion lms strategies in sensor networks with noisy input data," *IEEE/ACM Transactions on Networking*, vol. 24, no. 1, pp. 3–14, Feb 2016.
[18] M. O. Sayin and S. S. Kozat, "Single bit and reduced dimension diffusion strategies over distributed networks," *Signal Processing Letters, IEEE*, vol. 20, no. 10, pp. 976–979, 2013.
[19] ——, "Compressive diffusion strategies over distributed networks for reduced communication load," *Signal Processing, IEEE Transactions on*, vol. 62, no. 20, pp. 5308–5323, 2014.
[20] S. Chouvardas, K. Slavakis, and S. Theodoridis, "Trading off complexity with communication costs in distributed adaptive learning via Krylov subspaces for dimensionality reduction," *Selected Topics in Signal Processing, IEEE Journal of*, vol. 7, no. 2, pp. 257–273, 2013.
[21] R. Arablouei, S. Werner, Y.-F. Huang, and K. Dogancay, "Distributed least mean-square estimation with partial diffusion," *Signal Processing, IEEE Transactions on*, vol. 62, no. 2, pp. 472–484, 2014.
[22] R. Arablouei, K. Dogancay, S. Werner, and Y.-F. Huang, "Adaptive distributed estimation based on recursive least-squares and partial diffusion," *Signal Processing, IEEE Transactions on*, vol. 62, no. 14, pp. 3510–3522, 2014.
[23] J. R. Deller Jr and Y. F. Huang, "Set-membership identification and filtering for signal processing applications," *Circuits, Systems and Signal Processing*, vol. 21, no. 1, pp. 69–82, 2002.
[24] S. Gollamudi, S. Nagaraj, S. Kapoor, and Y.-F. Huang, "Set-membership filtering and a set-membership normalized LMS algorithm with an adaptive step size," *Signal Processing Letters, IEEE*, vol. 5, no. 5, pp. 111–114, 1998.
[25] J. R. Deller Jr, M. Nayeri, and S. F. Odeh, "Least-square identification with error bounds for real-time signal processing and control," *Proceedings of the IEEE*, vol. 81, no. 6, pp. 815–849, 1993.
[26] K. Dogancay, *Partial-update adaptive signal processing: Design Analysis and Implementation*. Academic Press, 2008.
[27] A. H. Sayed, "Adaptive Filters. Hoboken," 2008.
[28] M. Godavarti and A. O. Hero III, "Partial update LMS algorithms," *Signal Processing, IEEE Transactions on*, vol. 53, no. 7, pp. 2382–2399, 2005.
[29] S. C. Douglas, "Adaptive filters employing partial updates," *Circuits and Systems II: Analog and Digital Signal Processing, IEEE Transactions on*, vol. 44, no. 3, pp. 209–216, 1997.
[30] J. R. Treichler, C. R. Johnson, and M. G. Larimore, *Theory and design of adaptive filters*. Wiley, 1987.

Amir Rastegarnia completed his PhD degree in the electrical engineering at the University of Tabriz, Tabriz, Iran, in 2011. In 2011, he joined the Department of Electrical Engineering, Malayer University, as Assistant Professor. His current research interests are theory and methods for adaptive and statistical signal processing, distributed adaptive estimation, as well as signal processing for communications. He is a Member of IEEE.

Azam Khalili received the PhD degree in electrical engineering from the University of Tabriz, Tabriz, Iran, in 2011. In 2011, she joined the Department of Electrical Engineering, Malayer University, as Assistant Professor. Her current research interests are theory and methods for adaptive filtering, distributed adaptive estimation, as well as signal processing for communications. She is a Member of IEEE.

Wael M. Bazzi graduated from the American University of Beirut (AUB), Lebanon, in 1996. He received the M.E. degree from AUB in 1999 and the Ph.D. degree from the University of Waterloo, Canada, in 2001. He is currently an associate professor at the American University in Dubai. His research interests include wireless communication and networks, especially the optimization and modeling aspects of communication networks and systems.



Saeid Sanei (SM05) received his PhD in signal processing from Imperial College London, UK. He has been a member of academic staff in Iran, Singapore, and the UK. He has published three monograms, a number of book chapters, and over 320 papers in peer reviewed journals and conference proceedings. His research interest is in adaptive filtering, cooperative learning, multi-way, multimodal, and multichannel signal processing with applications to biomedical, audio, biometrics, and communication signals and images. He has served as an Associate Editor for the IEEE Signal Processing Letters, IEEE Signal Processing Magazine, and Journal of Computational Intelligence and Neuroscience.

Vahid Vahidpour received the M.Sc degree in communication engineering from Malayer University, Hamedan, Iran, in 2016. His research interests include underwater acoustic, distributed and adaptive signal processing, and biomedical signal processing. Mr. Vahidpour is a student member of the IEEE.