



**HAL**  
open science

## Diffusion LMS with Communication Delays: Stability and Performance Analysis

Fei Hua, Roula Nassif, Cédric Richard, Haiyan Wang, Ali H Sayed

► **To cite this version:**

Fei Hua, Roula Nassif, Cédric Richard, Haiyan Wang, Ali H Sayed. Diffusion LMS with Communication Delays: Stability and Performance Analysis. IEEE Signal Processing Letters, 2020, 27, pp.730 - 734. 10.1109/lsp.2020.2990086 . hal-03347250

**HAL Id: hal-03347250**

**<https://hal.science/hal-03347250>**

Submitted on 17 Sep 2021

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Diffusion LMS with Communication Delays: Stability and Performance Analysis

Fei Hua, Roula Nassif, *Member, IEEE*, Cédric Richard, *Senior Member, IEEE*,  
Haiyan Wang, and Ali H. Sayed, *Fellow, IEEE*

**Abstract**—We study the problem of distributed estimation over adaptive networks where communication delays exist between nodes. In particular, we investigate the diffusion Least-Mean-Square (LMS) strategy where delayed intermediate estimates (due to the communication channels) are employed during the combination step. One important question is: Do the delays affect the stability condition and performance? To answer this question, we conduct a detailed performance analysis in the mean and in the mean-square-error sense of the diffusion LMS with delayed estimates. Stability conditions, transient and steady-state mean-square-deviation (MSD) expressions are provided. One of the main findings is that diffusion LMS with delays can still converge under the same step-sizes condition of the diffusion LMS without delays. Finally, simulation results illustrate the theoretical findings.

**Index Terms**—Communication delays, distributed estimation, diffusion LMS, performance analysis, stability.

## I. INTRODUCTION

**D**ISTRIBUTED estimation aims to estimate a parameter vector of interest through local computation and cooperation among neighboring nodes. Distributed solutions are attractive in terms of robustness, scalability, and communication overheads. There has been extensive work in the literature studying distributed strategies, such as incremental strategies [1]–[3], consensus strategies [4]–[6], and diffusion strategies [7]–[9].

Most prior works on distributed estimation assume that the network is synchronous and that there is no noise or delay in communication. There exist several works studying distributed estimation in the presence of asynchronous events or imperfect information exchanges. For example, the performance

This work was supported in part by the NSFC under Grants 61901385, 61671386, 61771396, and in part by National Key R&D Program of China (2016YFC1400200). The work of C. Richard was supported in part by the ANR under Grant ANR-19-CE48-0002. The work of A. H. Sayed was supported in part by Grant 205121-184999 from the Swiss National Science Foundation.

F. Hua is with the School of Marine Science and Technology, Northwestern Polytechnical University, Xi’an 710072, China, and also with the Université Côte d’Azur, OCA, CNRS, Nice 06108, France (email: fei.hua@oca.eu).

R. Nassif is with the American University of Beirut, Beirut 1107 2020, Lebanon (email: roula.nassif@aub.edu.lb).

C. Richard is with the Université Côte d’Azur, OCA, CNRS, Nice 06108, France (email: cedric.richard@unice.fr).

H. Wang is with the School of Marine Science and Technology, Northwestern Polytechnical University, Xi’an, China, and also with the School of Electronic Information and Artificial Intelligence, Shaanxi University of Science and Technology, Xi’an 710021, China (email: hywang@sust.edu.cn).

A. H. Sayed is with the School of Engineering, Ecole Polytechnique Fédérale de Lausanne, Lausanne CH-1015, Switzerland (email: ali.sayed@epfl.ch).

of consensus and gossip strategies is investigated in [10]–[13] in the presence of link-failures, switching topology, and noisy links. Single-task and multi-task diffusion strategies over asynchronous networks are also examined in [14]–[17], while diffusion LMS over multi-task networks is investigated in [18] in the presence of noisy links.

The aforementioned works focus on asynchronous events such as link-failures, switching topology, noisy links, and agents turning on and off randomly. Another critical challenge in a distributed implementation is the presence of communication delays. For instance, in underwater acoustic sensor networks, one of the main characteristics of underwater communication channels is their long propagation delays due to the low speed of sound. It is therefore crucial to consider distributed solutions that take into account the communication delays. In [19], the authors studied a situation where the agents in the network seek average-consensus in the presence of time delays. It was shown that, in a *deterministic* setting, time delays should be smaller than a threshold to reach consensus. Recently, several other works have also derived interesting findings on consensus strategies in the presence of communication delays [20]–[25]. In [22], the authors propose and analyze the decentralized asynchronous primal-dual algorithm using fixed step-sizes which converges to the exact solution. In [23], a distributed consensus algorithm is proposed and analyzed for continuous-time multiagent systems in the presence of communication delays. Proportional-Integral (PI) consensus-based distributed optimization algorithm is presented to handle constant communication delays in [24]. In most cases, these works are problematic for processing streaming data in an *adaptive* context. It has been shown that diffusion strategies outperform consensus or primal-dual strategies in the *stochastic* setting due to gradient noise [26], [27]. By using *constant* step-sizes, diffusion strategies ensure continuous learning and adaptation. In this letter, we carry out a detailed performance analysis of diffusion LMS strategy in the presence of delays and provide stability conditions in the mean and mean-square sense. One of the main findings of this work is that the communication delays do not affect the convergence condition of adaptive diffusion networks.

## II. DATA MODEL AND DIFFUSION STRATEGIES

Consider a network of  $N$  agents, labeled with  $k = 1, \dots, N$ . At each time instant  $i \geq 0$ , each agent  $k$  is assumed to have access to a zero-mean scalar measurement  $d_{k,i} \in \mathbb{R}$  and a real-valued regression vector  $\mathbf{x}_{k,i} \in \mathbb{R}^M$  with positive-definite covariance matrix  $\mathbf{R}_k = \mathbb{E}\{\mathbf{x}_{k,i}\mathbf{x}_{k,i}^\top\}$ . Let  $\mathbf{r}_{dx,k} \triangleq \mathbb{E}\{d_{k,i}\mathbf{x}_{k,i}\}$ .

The data  $\{d_{k,i}, \mathbf{x}_{k,i}\}$  are assumed to be related via the linear regression model:

$$d_{k,i} = \mathbf{x}_{k,i}^\top \mathbf{w}^* + v_{k,i} \quad (1)$$

where  $\mathbf{w}^* \in \mathbb{R}^M$  is an unknown parameter vector to be estimated, and  $v_{k,i} \in \mathbb{R}$  is a zero-mean spatially independent measurement noise with variance  $\sigma_{v,k}^2$ . In order to find an estimate for  $\mathbf{w}^*$ , the objective of the network is to minimize the cost function  $J(\mathbf{w}) : \mathbb{R}^M \rightarrow \mathbb{R}$  given by:

$$\min_{\mathbf{w}} J(\mathbf{w}) = \sum_{k=1}^N \mathbb{E}(d_{k,i} - \mathbf{x}_{k,i}^\top \mathbf{w})^2. \quad (2)$$

It can be verified that the solution of the above problem is given by  $\mathbf{w}^* = (\sum_{k=1}^N \mathbf{R}_k)^{-1} (\sum_{k=1}^N \mathbf{r}_{dx,k})$ . This requires the signal statistical information  $\mathbf{R}_k$  and  $\mathbf{r}_{dx,k}$ , which are rarely available in practice. To solve the problem in a fully-distributed and adaptive manner, the following adapt-then-combine (ATC) diffusion LMS strategy can be employed [7]:

$$\psi_{k,i} = \mathbf{w}_{k,i-1} + \mu_k \mathbf{x}_{k,i} (d_{k,i} - \mathbf{x}_{k,i}^\top \mathbf{w}_{k,i-1}) \quad (3a)$$

$$\mathbf{w}_{k,i} = \sum_{\ell \in \mathcal{N}_k} a_{\ell k} \psi_{\ell,i} \quad (3b)$$

where  $\mathcal{N}_k$  denotes the set of neighbors of agent  $k$  including itself and  $\mathbf{w}_{k,i}$  denotes the estimate of  $\mathbf{w}^*$  at agent  $k$  and iteration  $i$ . The first step (3a) is an adaptation step where agent  $k$  uses its own data available at time  $i$  to update the previous estimate  $\mathbf{w}_{k,i-1}$  to an intermediate estimate  $\psi_{k,i}$ . Then, in the combination step (3b), agent  $k$  convexly combines the intermediate estimates  $\psi_{\ell,i}$  from its neighbors to obtain  $\mathbf{w}_{k,i}$ . The parameter  $\mu_k$  is a positive step-size, and the combination coefficients  $\{a_{\ell k}\}$  are non-negative, chosen to satisfy the following conditions:

$$\sum_{\ell=1}^N a_{\ell k} = 1, \text{ and } \begin{cases} a_{\ell k} > 0, & \text{if } \ell \in \mathcal{N}_k, \\ a_{\ell k} = 0, & \text{otherwise.} \end{cases} \quad (4)$$

The above conditions imply that the matrix  $[\mathbf{A}]_{\ell k} = a_{\ell k}$  collecting the parameters  $a_{\ell k}$  is left-stochastic.

Notice that, in the combination step (3b), each node  $k$  at time  $i$  is assumed to have access to the estimates  $\psi_{\ell,i}$  obtained by its neighbors at the *same* time instant  $i$ . This requires the network to be time synchronized and each node should transmit its estimate to its neighbors before the next iteration  $i + 1$ . In this work, we are interested in scenarios where there are time delays in information exchange. In this case, at time  $i$ , the estimates  $\psi_{\ell,i}$  from neighbors are not available at node  $k$ . Instead, previous estimates may just have been received. In principle, each node  $k$  could pause the adaptation step until receiving the required timely estimates and delay the processing of data by the network. Alternatively, we can directly use the delayed information without any additional complexity. By doing so, we arrive at ATC diffusion LMS with delays:

$$\psi_{k,i} = \mathbf{w}_{k,i-1} + \mu_k \mathbf{x}_{k,i} (d_{k,i} - \mathbf{x}_{k,i}^\top \mathbf{w}_{k,i-1}) \quad (5a)$$

$$\mathbf{w}_{k,i} = \sum_{\ell \in \mathcal{N}_k} a_{\ell k} \psi_{\ell,i-\tau_{\ell k}} \quad (5b)$$

where the adaptation step (5a) is the same as (3a). However, instead of using the timely estimates  $\psi_{\ell,i}$ , the combination step (5b) uses delayed estimates  $\psi_{\ell,i-\tau_{\ell k}}$  from neighbors, where  $\tau_{\ell k} \geq 0$  is an integer denoting the communication delay from node  $\ell$  to  $k$ .

### III. PERFORMANCE ANALYSIS

We now analyze the stability and performance of diffusion LMS with delays (5). Before proceeding, we introduce the following independence assumption.

**Assumption 1** (Independent regressors). *The regression vector  $\mathbf{x}_{k,i}$  arises from a stationary random process that is temporally white and spatially independent with covariance matrix  $\mathbf{R}_k = \mathbb{E}\{\mathbf{x}_{k,i} \mathbf{x}_{k,i}^\top\} > 0$ .* ■

A consequence of Assumption 1 is that we can consider the regressors  $\{\mathbf{x}_{k,i}\}$  independent of  $\mathbf{w}_{\ell,j}$  for all  $\ell$  and  $j < i$ . Although not true in general, this assumption is commonly employed for analyzing adaptive filters and networks since it simplifies the derivations without constraining the conclusions. Furthermore, there are extensive results in the adaptive filtering literature indicating that the performance results obtained using this assumption match well the actual performance for sufficiently small step-sizes [28].

#### A. Error Recursion

We introduce the error vectors at node  $k$  and time instant  $i$ :

$$\tilde{\psi}_{k,i} \triangleq \mathbf{w}^* - \psi_{k,i}, \quad \tilde{\mathbf{w}}_{k,i} \triangleq \mathbf{w}^* - \mathbf{w}_{k,i}, \quad (6)$$

and collect all error vectors into network block error vectors:

$$\tilde{\psi}_i \triangleq \text{col}\{\tilde{\psi}_{1,i}, \dots, \tilde{\psi}_{N,i}\}, \quad \tilde{\mathbf{w}}_i \triangleq \text{col}\{\tilde{\mathbf{w}}_{1,i}, \dots, \tilde{\mathbf{w}}_{N,i}\}. \quad (7)$$

Subtracting  $\mathbf{w}^*$  from both sides of the adaptation step (5a) and using the data model (1), it can be verified that

$$\tilde{\psi}_i = (\mathbf{I}_{MN} - \mathcal{M}\mathcal{R}_i) \tilde{\mathbf{w}}_{i-1} - \mathcal{M}\mathbf{s}_i \quad (8)$$

where  $\mathcal{M}$  and  $\mathcal{R}_i$  are  $N \times N$  block diagonal matrices with each block of size  $M \times M$ ,  $\mathbf{s}_i$  is an  $N \times 1$  block vector whose entries are of size  $M \times 1$  each:

$$\mathcal{M} \triangleq \text{bdiag}\{\mu_1 \mathbf{I}_M, \dots, \mu_N \mathbf{I}_M\}, \quad (9)$$

$$\mathcal{R}_i \triangleq \text{bdiag}\{\mathbf{x}_{1,i} \mathbf{x}_{1,i}^\top, \dots, \mathbf{x}_{N,i} \mathbf{x}_{N,i}^\top\}, \quad (10)$$

$$\mathbf{s}_i \triangleq \text{col}\{\mathbf{x}_{1,i} v_{1,i}, \dots, \mathbf{x}_{N,i} v_{N,i}\}. \quad (11)$$

It holds that  $\mathcal{R} = \mathbb{E}\mathcal{R}_i = \text{bdiag}\{\mathbf{R}_1, \dots, \mathbf{R}_N\}$  and  $\mathbb{E}\mathbf{s}_i = \mathbf{0}$ .

Due to the communication delays, the relation between  $\tilde{\mathbf{w}}_i$  and  $\tilde{\psi}_i$  cannot be obtained by simply subtracting  $\mathbf{w}^*$  from both sides of the combination step (5b). Following the same line of reasoning as in [29], [30], we introduce the following  $N(\Gamma + 1) \times 1$  extended network block error vectors with each block of size  $M \times 1$ :

$$\tilde{\psi}_i^e \triangleq \text{col}\{\tilde{\psi}_{1,i}, \dots, \tilde{\psi}_{N,i}, \tilde{\psi}_{1,i-1}, \dots, \tilde{\psi}_{N,i-1}, \dots, \tilde{\psi}_{1,i-\Gamma}, \dots, \tilde{\psi}_{N,i-\Gamma}\}, \quad (12)$$

$$\tilde{\mathbf{w}}_i^e \triangleq \text{col}\{\tilde{\mathbf{w}}_{1,i}, \dots, \tilde{\mathbf{w}}_{N,i}, \tilde{\psi}_{1,i}, \dots, \tilde{\psi}_{N,i}, \dots, \tilde{\psi}_{1,i-\Gamma+1}, \dots, \tilde{\psi}_{N,i-\Gamma+1}\}, \quad (13)$$

where  $\Gamma = \max\{\tau_{\ell k}\}, 1 \leq \ell, k \leq N$ . For simplicity, we let  $T \triangleq \Gamma + 1$ . Using the fact that the matrix  $\mathbf{A}$  is left-stochastic, and from (5b), we obtain

$$\tilde{\mathbf{w}}_i^e = \mathbf{A}^e \tilde{\boldsymbol{\psi}}_i^e \quad (14)$$

with  $\mathbf{A}^e$  an  $MNT \times MNT$  matrix given by

$$\mathbf{A}^e = \begin{bmatrix} \mathbf{A}_0^\top & \mathbf{A}_1^\top & \cdots & \mathbf{A}_\Gamma^\top \\ & \mathbf{I}_{MNT} & & \mathbf{0}_{(MNT) \times (MN)} \end{bmatrix} \quad (15)$$

where  $\mathbf{A}_\tau \triangleq \mathbf{A}_\tau \otimes \mathbf{I}_M$  ( $\tau = 0, 1, \dots, \Gamma$ ) with the  $(\ell, k)$ -th entry of the  $N \times N$  matrix  $\mathbf{A}_\tau$  is given by

$$[\mathbf{A}_\tau]_{\ell k} = [\mathbf{A}]_{\ell k}, \text{ if } \tau_{\ell k} = \tau; \quad [\mathbf{A}_\tau]_{\ell k} = 0, \text{ otherwise.} \quad (16)$$

Observe that  $\mathbf{A} = \sum_{\tau=0}^{\Gamma} \mathbf{A}_\tau$  and  $\mathbf{A} \triangleq \mathbf{A} \otimes \mathbf{I}_M = \sum_{\tau=0}^{\Gamma} \mathbf{A}_\tau$ . Likewise, from (8), we can derive the extended error vector recursive relation between  $\tilde{\boldsymbol{\psi}}_i^e$  and  $\tilde{\mathbf{w}}_{i-1}^e$ :

$$\tilde{\boldsymbol{\psi}}_i^e = (\mathbf{I}_{MNT} - \mathbf{R}_i^e) \tilde{\mathbf{w}}_{i-1}^e - \mathbf{s}_i^e \quad (17)$$

with

$$\mathbf{R}_i^e \triangleq \begin{bmatrix} \mathcal{M}\mathcal{R}_i & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}, \quad \mathbf{s}_i^e \triangleq \begin{bmatrix} \mathcal{M}\mathbf{s}_i \\ \mathbf{0}_{(MNT) \times 1} \end{bmatrix}. \quad (18)$$

By combining (14) and (17), we conclude that the extended network error  $\tilde{\mathbf{w}}_i^e$  evolves according to the following recursion:

$$\tilde{\mathbf{w}}_i^e = \mathbf{B}_i \tilde{\mathbf{w}}_{i-1}^e - \mathbf{A}^e \mathbf{s}_i^e \quad (19)$$

where  $\mathbf{B}_i \triangleq \mathbf{A}^e (\mathbf{I}_{MNT} - \mathbf{R}_i^e)$ .

### B. Mean-error Behavior

Taking the expectation of both sides of (19), using Assumption 1 and the fact that  $\mathbb{E}\mathbf{s}_i = \mathbf{0}$ , we arrive at the mean-error recursion:

$$\mathbb{E}\tilde{\mathbf{w}}_i^e = \mathbf{A}^e (\mathbf{I}_{MNT} - \mathbf{R}^e) \mathbb{E}\tilde{\mathbf{w}}_{i-1}^e = \mathbf{B} \mathbb{E}\tilde{\mathbf{w}}_{i-1}^e \quad (20)$$

$$\mathbf{R}^e = \mathbb{E}\mathbf{R}_i^e = \text{bdiag}\{\mathcal{M}\mathcal{R}, \mathbf{0}, \dots, \mathbf{0}\}, \quad (21)$$

$$\mathbf{B} = \mathbf{A}^e (\mathbf{I}_{MNT} - \mathbf{R}^e). \quad (22)$$

**Lemma 1.** Consider the block matrix  $\mathbf{B}$  defined by (22). The matrix  $\mathbf{B}$  is stable, i.e., its spectral radius  $\rho(\mathbf{B})$  is less than 1, when  $\|\mathbf{I}_{MN} - \mathcal{M}\mathcal{R}\|_{b,\infty} < 1$  where the notation  $\|\cdot\|_{b,\infty}$  denotes the block maximum norm of its argument.<sup>1</sup>

*Proof:* The proof can be found in the supplemental material. ■

**Theorem 1** (Convergence in the mean). Assume the linear data model (1) and Assumption 1 hold. Then, for any initial condition, algorithm (5) converges asymptotically in the mean toward the optimal vector  $\mathbf{w}^*$  if the step-sizes in  $\mathcal{M}$  are chosen to satisfy:

$$0 < \mu_k < \frac{2}{\lambda_{\max}(\mathbf{R}_k)}, \quad k = 1, \dots, N. \quad (23)$$

*Proof:* The proof can be found in the supplemental material. ■

Theorem 1 allows us to conclude that diffusion LMS with communication delays will continue to converge in the mean

sense under the same step-sizes condition of the algorithm without delays. It is worth noting that this work focuses on single-task problems where all nodes seek to estimate the same parameter vector  $\mathbf{w}^*$  in (1). In this case, the resulting estimates  $\mathbf{w}_{k,i}$  will be unbiased according to Theorem 1. When heterogeneity in the model exists, diffusion multitask algorithms [32], [33] and exact diffusion algorithms [34] can be used in order to obtain unbiased estimates.

### C. Variance Relation

We now study the mean-square-error behavior. We consider the mean-square error vector weighted by a positive semi-definite matrix  $\boldsymbol{\Sigma}$ , i.e.,  $\mathbb{E}\|\tilde{\mathbf{w}}_i^e\|_{\boldsymbol{\Sigma}}^2 \triangleq \mathbb{E}(\tilde{\mathbf{w}}_i^e)^\top \boldsymbol{\Sigma} \tilde{\mathbf{w}}_i^e$ . The freedom in selecting  $\boldsymbol{\Sigma}$  allow us to derive different performance measures about the network and the nodes. Evaluating the weighted square measures on both sides of (19), we get:

$$\begin{aligned} \|\tilde{\mathbf{w}}_i^e\|_{\boldsymbol{\Sigma}}^2 &= \|\mathbf{B}_i \tilde{\mathbf{w}}_{i-1}^e - \mathbf{A}^e \mathbf{s}_i^e\|_{\boldsymbol{\Sigma}}^2 \\ &= (\tilde{\mathbf{w}}_{i-1}^e)^\top \mathbf{B}_i^\top \boldsymbol{\Sigma} \mathbf{B}_i \tilde{\mathbf{w}}_{i-1}^e + (\mathbf{A}^e \mathbf{s}_i^e)^\top \boldsymbol{\Sigma} \mathbf{A}^e \mathbf{s}_i^e - \\ &\quad (\mathbf{B}_i \tilde{\mathbf{w}}_{i-1}^e)^\top \boldsymbol{\Sigma} \mathbf{A}^e \mathbf{s}_i^e - (\mathbf{A}^e \mathbf{s}_i^e)^\top \boldsymbol{\Sigma} \mathbf{B}_i \tilde{\mathbf{w}}_{i-1}^e. \end{aligned} \quad (24)$$

Taking the expectation of both sides of (24), using Assumption 1 and the fact that the expectations of the last two terms on the right-hand side (RHS) of (24) are zero, we obtain:

$$\mathbb{E}\|\tilde{\mathbf{w}}_i^e\|_{\boldsymbol{\Sigma}}^2 = \mathbb{E}\|\tilde{\mathbf{w}}_{i-1}^e\|_{\boldsymbol{\Sigma}'}^2 + \mathbb{E}\{(\mathbf{A}^e \mathbf{s}_i^e)^\top \boldsymbol{\Sigma} \mathbf{A}^e \mathbf{s}_i^e\} \quad (25)$$

where  $\boldsymbol{\Sigma}' \triangleq \mathbb{E}\{\mathbf{B}_i^\top \boldsymbol{\Sigma} \mathbf{B}_i\}$ . Let  $\boldsymbol{\sigma} \triangleq \text{vec}(\boldsymbol{\Sigma})$  denote the vector obtained by stacking the columns of the matrix  $\boldsymbol{\Sigma}$  on top of each other. Note that, in the sequel, we will use the notation  $\|\cdot\|_{\boldsymbol{\Sigma}'}^2$  and  $\|\cdot\|_{\boldsymbol{\Sigma}}^2$  interchangeably to refer to the same quantity. Considering the following properties of matrices:

$$\text{vec}(\mathbf{A}\boldsymbol{\Sigma}\mathbf{B}) = (\mathbf{B}^\top \otimes \mathbf{A})\boldsymbol{\sigma}, \quad (26)$$

$$\text{Tr}(\boldsymbol{\Sigma}\mathbf{B}) = [\text{vec}(\mathbf{B}^\top)]^\top \boldsymbol{\sigma}, \quad (27)$$

we find that  $\boldsymbol{\sigma}' \triangleq \text{vec}(\boldsymbol{\Sigma}') = \mathbb{E}\{\mathbf{B}_i^\top \otimes \mathbf{B}_i^\top\}\boldsymbol{\sigma}$ . Let  $\mathcal{F} \triangleq \mathbb{E}\{\mathbf{B}_i^\top \otimes \mathbf{B}_i^\top\}$ . Under sufficiently small step-sizes [28], the matrix  $\mathcal{F}$  can be approximated by:

$$\mathcal{F} \approx \mathbf{B}^\top \otimes \mathbf{B}^\top. \quad (28)$$

The first item on the RHS of (25) can be rewritten as  $\mathbb{E}\|\tilde{\mathbf{w}}_{i-1}^e\|_{\boldsymbol{\Sigma}'}^2 = \mathbb{E}\|\tilde{\mathbf{w}}_{i-1}^e\|_{\mathcal{F}\boldsymbol{\sigma}}^2$ . Now we evaluate the second term:

$$\mathbb{E}\{(\mathbf{A}^e \mathbf{s}_i^e)^\top \boldsymbol{\Sigma} \mathbf{A}^e \mathbf{s}_i^e\} = \text{Tr}(\boldsymbol{\Sigma}\mathcal{G}) = [\text{vec}(\mathcal{G}^\top)]^\top \boldsymbol{\sigma} \quad (29)$$

where

$$\mathcal{G} \triangleq \mathbf{A}^e \mathcal{S}^e (\mathbf{A}^e)^\top, \quad (30)$$

$$\mathcal{S}^e \triangleq \mathbb{E}\{\mathbf{s}_i^e (\mathbf{s}_i^e)^\top\} = \text{bdiag}\{\mathcal{S}, \mathbf{0}, \dots, \mathbf{0}\}, \quad (31)$$

$$\mathcal{M} \triangleq \mathcal{M} \cdot \text{bdiag}\{\sigma_{v,k}^2 \mathbf{R}_k\}_{k=1}^N \cdot \mathcal{M}. \quad (32)$$

Therefore, the variance relation (25) can be approximated as

$$\mathbb{E}\|\tilde{\mathbf{w}}_i^e\|_{\boldsymbol{\sigma}}^2 = \mathbb{E}\|\tilde{\mathbf{w}}_{i-1}^e\|_{\mathcal{F}\boldsymbol{\sigma}}^2 + [\text{vec}(\mathcal{G}^\top)]^\top \boldsymbol{\sigma}. \quad (33)$$

**Theorem 2** (Mean-square stability). Consider the same settings as in Theorem 1. The diffusion LMS with delays algorithm (5) is mean-square stable if the matrix  $\mathcal{F}$  is stable. Assuming further that the step-sizes are small enough to justify (28), condition (23) ensures mean-square stability.

*Proof:* The proof can be found in the supplemental material. ■

<sup>1</sup>For more details and properties of the block maximum norm, refer to [31, Appendix D].

#### D. Network Transient and Steady-state MSD

Iterating (33) starting from  $i = 0$ , we obtain:

$$\mathbb{E}\|\tilde{\mathbf{w}}_i^e\|_{\sigma}^2 = \mathbb{E}\|\tilde{\mathbf{w}}_{-1}^e\|_{\mathcal{F}^{i+1}\sigma}^2 + [\text{vec}(\mathcal{G}^{\top})]^{\top} \sum_{t=0}^i \mathcal{F}^t \sigma \quad (34)$$

where  $\tilde{\mathbf{w}}_{-1}^e = \mathbf{1}_{NT} \otimes \mathbf{w}^*$  is an initial condition by assuming  $\mathbf{w}_{-1} = \mathbf{0}$ . Comparing relation (34) at time instants  $i$  and  $i-1$ , we can derive the weighted variance recursion

$$\mathbb{E}\|\tilde{\mathbf{w}}_i^e\|_{\sigma}^2 = \mathbb{E}\|\tilde{\mathbf{w}}_{i-1}^e\|_{\sigma}^2 + \|\tilde{\mathbf{w}}_{-1}^e\|_{(\mathcal{F}-I)\mathcal{F}^i\sigma}^2 + [\text{vec}(\mathcal{G}^{\top})]^{\top} \mathcal{F}^i \sigma. \quad (35)$$

Let  $\zeta_i \triangleq \frac{1}{N} \mathbb{E}\|\tilde{\mathbf{w}}_i\|^2$  denote the network transient MSD averaged over all nodes at time  $i$ . By replacing  $\sigma$  with  $\bar{\sigma} = \text{vec}(\bar{\Sigma})$ ,  $\bar{\Sigma} = \text{bdiag}\{\mathbf{I}_{MN}, \mathbf{0}, \dots, \mathbf{0}\}$ , we find that the network transient MSD evolves according to

$$\zeta_i = \zeta_{i-1} + \frac{1}{N} \left( \|\tilde{\mathbf{w}}_{-1}^e\|_{(\mathcal{F}-I)\mathcal{F}^i\bar{\sigma}}^2 + [\text{vec}(\mathcal{G}^{\top})]^{\top} \mathcal{F}^i \bar{\sigma} \right) \quad (36)$$

with  $\zeta_{-1} = \frac{1}{N} (\tilde{\mathbf{w}}_{-1}^e)^{\top} \bar{\Sigma} \tilde{\mathbf{w}}_{-1}^e$ . Notice that the evaluation of (36) involves the manipulation of the  $(MNT)^2 \times (MNT)^2$  matrix, which will be prohibitive in computing. However, using property (27), expression (36) can be rewritten as

$$\zeta_i = \zeta_{i-1} + \frac{1}{N} \text{Tr} \left( \tilde{\mathbf{w}}_{-1}^e (\tilde{\mathbf{w}}_{-1}^e)^{\top} ((\mathcal{B}^{i+1})^{\top} \bar{\Sigma} \mathcal{B}^{i+1} - (\mathcal{B}^i)^{\top} \bar{\Sigma} \mathcal{B}^i) + (\mathcal{B}^i)^{\top} \bar{\Sigma} \mathcal{B}^i \mathcal{G} \right) \quad (37)$$

where the matrix operations ease to the order of  $\mathcal{O}(MNT)$ .

The network steady-state MSD is defined as  $\zeta^* \triangleq \frac{1}{N} \lim_{i \rightarrow \infty} \mathbb{E}\|\tilde{\mathbf{w}}_i\|^2$ . Assuming that  $\mathcal{F}$  is stable and that the algorithm converges in the mean-square sense, taking the limit on both sides of (34), we observe that the first term on the RHS converges to zero. Then, by setting  $\sigma = \frac{1}{N} \text{vec}(\bar{\Sigma})$  and using properties (26) and (27), we get:

$$\zeta^* = \frac{1}{N} \sum_{t=0}^{\infty} \text{Tr} \left( (\mathcal{B}^t)^{\top} \bar{\Sigma} \mathcal{B}^t \mathcal{G} \right). \quad (38)$$

#### IV. SIMULATIONS

We consider a network of 30 nodes with the topology depicted in Fig. 1(a). The length of the parameter vector is set to  $M = 10$  and the optimal vector is  $\mathbf{w}^* = [0.495, -0.134, 0.139, -0.328, 0.367, -0.049, -0.141, -1.858, -0.253, -0.602]^{\top}$ . The regression vectors  $\mathbf{x}_{k,i}$  are generated from a zero-mean Gaussian distribution with covariance matrix  $\mathbf{R}_k = \sigma_{x,k}^2 \mathbf{I}_3$ . The noises  $v_{k,i}$  are zero-mean i.i.d. Gaussian random variables with variances  $\sigma_{v,k}^2$ . The variances  $\sigma_{x,k}^2$  and  $\sigma_{v,k}^2$  are shown in Fig. 1(b). The communication delay  $\tau_{\ell k}$  between two connected nodes  $\ell$  and  $k$  is proportional to their distance. The step-sizes are set to  $\mu_k = \mu$  for all nodes. Combination coefficients for diffusion strategies are chosen according to the uniform rule, i.e.,  $a_{\ell k} = 1/|\mathcal{N}_k|$  for  $\ell \in \mathcal{N}_k$ . All simulated results are averaged over 500 independent trials.

In the first experiment, we set  $\mu = 0.02$  for non-cooperative LMS, diffusion LMS with delays, diffusion LMS with ideal communications where there is no delay, and synchronous diffusion LMS. Note that, for the synchronous diffusion LMS,

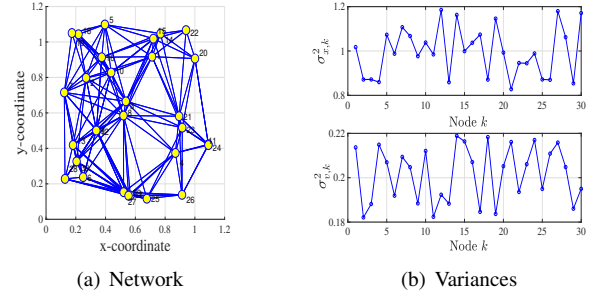


Fig. 1. Simulation settings. Variances are generated as  $\sigma_{x,k}^2 \sim \mathcal{U}(0.8, 1.2)$ ,  $\sigma_{v,k}^2 \sim \mathcal{U}(0.18, 0.22)$ .

all nodes need to wait for the longest delayed information to complete one adaptation and combination process. Observe from Fig. 2 that the diffusion LMS strategies perform better than non-cooperative LMS in terms of steady-state MSD, and that diffusion LMS with delays achieves better network steady-state MSD compared to synchronous diffusion LMS at a faster convergence rate and to diffusion LMS with ideal communications at a slower rate under the same step-sizes. In the second experiment, for comparison purposes, we set  $\mu = 0.035$  for diffusion LMS with delays in order to meet the same steady-state MSD. It is seen that diffusion LMS with delays converges much faster than the synchronous counterpart. This implies that one can adjust the step-sizes for diffusion LMS with delays to obtain faster convergence rate without additional asynchronous, computational and storage overheads. Moreover, the algorithm will be stable as long as the step-sizes are small enough, which is independent of delays. Finally, from Fig. 2, we observe that the simulated results match well the theoretical curves.

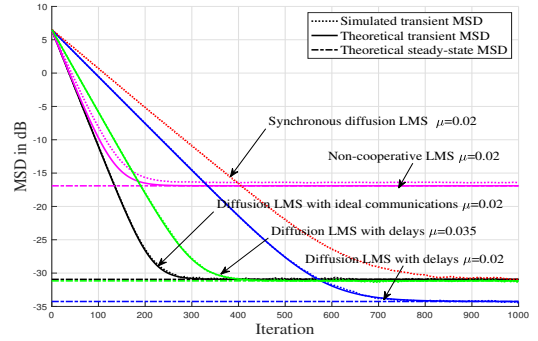


Fig. 2. MSD comparison.

#### V. CONCLUSION

In this work, we considered the problem of distributed estimation over adaptive networks in the presence of communication delays. We derived the stability condition for diffusion LMS strategy with delays. Stochastic behaviors in the mean and mean-square sense were also provided. Simulation results confirmed the theoretical findings.

## REFERENCES

- [1] A. Nedic and D. P. Bertsekas, "Incremental subgradient methods for nondifferentiable optimization," *SIAM J. Optim.*, vol. 12, no. 1, pp. 109–138, 2001.
- [2] D. Blatt, A. O. Hero, and H. Gauchman, "A convergent incremental gradient method with a constant step size," *SIAM J. Optim.*, vol. 18, no. 1, pp. 29–51, 2007.
- [3] M. G. Rabbat and R. D. Nowak, "Quantized incremental algorithms for distributed optimization," *IEEE J. Sel. Areas Commun.*, vol. 23, no. 4, pp. 798–808, Apr. 2005.
- [4] L. Xiao, S. Boyd, and S. Lall, "A scheme for robust distributed sensor fusion based on average consensus," in *Proc. 2005 4th Int. Symp. Inf. Process. Sensor Netw.*, Boise, ID, USA, 2005, pp. 63–70.
- [5] R. Olfati-Saber, J. A. Fax, and R. M. Murray, "Consensus and cooperation in networked multi-agent systems," *Proc. IEEE*, vol. 95, no. 1, pp. 215–233, Jan. 2007.
- [6] P. Braca, S. Marano, and V. Matta, "Enforcing consensus while monitoring the environment in wireless sensor networks," *IEEE Trans. Signal Process.*, vol. 56, no. 7, pp. 3375–3380, Jun. 2008.
- [7] F. S. Cattivelli and A. H. Sayed, "Diffusion LMS strategies for distributed estimation," *IEEE Trans. Signal Process.*, vol. 58, no. 3, pp. 1035–1048, Mar. 2010.
- [8] A. H. Sayed, S.-Y. Tu, J. Chen, X. Zhao, and Z. J. Towfic, "Diffusion strategies for adaptation and learning over networks," *IEEE Signal Process. Mag.*, vol. 30, no. 3, pp. 155–171, May. 2013.
- [9] A. H. Sayed, "Adaptive networks," *Proc. IEEE*, vol. 102, no. 4, pp. 460–497, Apr. 2014.
- [10] J. Tsitsiklis, D. Bertsekas, and M. Athans, "Distributed asynchronous deterministic and stochastic gradient optimization algorithms," *IEEE Trans. Autom. Control*, vol. 31, no. 9, pp. 803–812, Sep. 1986.
- [11] S. Kar and J. M. F. Moura, "Distributed consensus algorithms in sensor networks with imperfect communication: Link failures and channel noise," *IEEE Trans. Signal Process.*, vol. 57, no. 1, pp. 355–369, Jan. 2009.
- [12] T. C. Aysal, M. E. Yildiz, A. D. Sarwate, and A. Scaglione, "Broadcast gossip algorithms for consensus," *IEEE Trans. Signal Process.*, vol. 57, no. 7, pp. 2748–2761, Jul. 2009.
- [13] S. Kar and J. M. F. Moura, "Distributed consensus algorithms in sensor networks: Quantized data and random link failures," *IEEE Trans. Signal Process.*, vol. 58, no. 3, pp. 1383–1400, Mar. 2010.
- [14] X. Zhao and A. H. Sayed, "Asynchronous adaptation and learning over networks—Part I: Modeling and stability analysis," *IEEE Trans. Signal Process.*, vol. 63, no. 4, pp. 811–826, Feb. 2015.
- [15] —, "Asynchronous adaptation and learning over networks—Part II: Performance analysis," *IEEE Trans. Signal Process.*, vol. 63, no. 4, pp. 827–842, Feb. 2015.
- [16] —, "Asynchronous adaptation and learning over networks—Part III: Comparison analysis," *IEEE Trans. Signal Process.*, vol. 63, no. 4, pp. 843–858, Feb. 2015.
- [17] R. Nassif, C. Richard, A. Ferrari, and A. H. Sayed, "Multitask diffusion adaptation over asynchronous networks," *IEEE Trans. Signal Process.*, vol. 64, no. 11, pp. 2835–2850, Jun. 2016.
- [18] R. Nassif, C. Richard, J. Chen, A. Ferrari, and A. H. Sayed, "Diffusion LMS over multitask networks with noisy links," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, Shanghai, China, May 2016, pp. 4583–4587.
- [19] R. Olfati-Saber and R. Murray, "Consensus problems in networks of agents with switching topology and time-delays," *IEEE Trans. Autom. Control*, vol. 49, no. 9, pp. 1520–1533, Sep. 2004.
- [20] K. I. Tsianos and M. G. Rabbat, "Distributed consensus and optimization under communication delays," in *Proc. 49th Annu. Allerton Conf. Commun. Control Comput.*, Monticello, IL, USA, Sep. 2011, pp. 974–982.
- [21] H. Wang, X. Liao, T. Huang, and C. Li, "Cooperative distributed optimization in multiagent networks with delays," *IEEE Trans. Syst., Man, Cybern., Syst.*, vol. 45, no. 2, pp. 363–369, Feb. 2015.
- [22] T. Wu, K. Yuan, Q. Ling, W. Yin, and A. H. Sayed, "Decentralized consensus optimization with asynchrony and delays," *IEEE Trans. Signal Inf. Process. Netw.*, vol. 4, no. 2, pp. 293–307, Jun. 2018.
- [23] S. Yang, Q. Liu, and J. Wang, "Distributed optimization based on a multiagent system in the presence of communication delays," *IEEE Trans. Syst., Man, Cybern., Syst.*, vol. 47, no. 5, pp. 717–728, Mar. 2017.
- [24] T. Hatanaka, N. Chopra, T. Ishizaki, and N. Li, "Passivity-based distributed optimization with communication delays using PI consensus algorithm," *IEEE Trans. Autom. Control*, vol. 63, no. 12, pp. 4421–4428, Apr. 2018.
- [25] P. Yi and L. Pavel, "Asynchronous distributed algorithms for seeking generalized nash equilibria under full and partial-decision information," *IEEE Trans. Cybern.*, pp. 1–13, to be published, 2019.
- [26] S.-Y. Tu and A. H. Sayed, "Diffusion strategies outperform consensus strategies for distributed estimation over adaptive networks," *IEEE Trans. Signal Process.*, vol. 60, no. 12, pp. 6217–6234, Dec. 2012.
- [27] Z. J. Towfic and A. H. Sayed, "Stability and performance limits of adaptive primal-dual networks," *IEEE Trans. Signal Process.*, vol. 63, no. 11, pp. 2888–2903, Jun. 2015.
- [28] A. H. Sayed, *Adaptive Filters*. Hoboken, NY, USA: John Wiley & Sons, 2008.
- [29] J.-W. Lee, S.-E. Kim, W.-J. Song, and A. H. Sayed, "Spatio-temporal diffusion strategies for estimation and detection over networks," *IEEE Trans. Signal Process.*, vol. 60, no. 8, pp. 4017–4034, Aug. 2012.
- [30] F. Hua, R. Nassif, C. Richard, and H. Wang, "Penalty-based multitask estimation with non-local linear equality constraints," in *Proc. IEEE Int. Workshop Comput. Adv. Multi-Sensor Adapt. Process.*, Curacao, Netherlands Antilles, Dec. 2017, pp. 1–5.
- [31] A. H. Sayed, "Diffusion adaptation over networks," in *Academic Press Library in Signal Processing*, S. Theodoridis and R. Chellappa, Eds. Academic Press, Elsevier, 2014, vol. 3, pp. 322–454.
- [32] R. Nassif, S. Vlaski, C. Richard, J. Chen, and A. H. Sayed, "Multitask learning over graphs," to appear in *IEEE Signal Process. Mag.* Also available as arXiv:2001.02112, May 2020.
- [33] J. Chen, C. Richard, and A. H. Sayed, "Multitask diffusion adaptation over networks," *IEEE Trans. Signal Process.*, vol. 62, no. 16, pp. 4129–4144, Aug. 2014.
- [34] K. Yuan, B. Ying, X. Zhao, and A. H. Sayed, "Exact diffusion for distributed optimization and learning—Part I: Algorithm development," *IEEE Trans. Signal Process.*, vol. 67, no. 3, pp. 708–723, Feb. 2019.