Performance of Clustered Multitask Diffusion LMS Suffering from Inter-Node Communication Delays

Vinay Chakravarthi Gogineni, *Member, IEEE*, Sayed Pouria Talebi and Stefan Werner, *Senior Member, IEEE*

***Abstract*—This paper studies a clustered multitask diffusion least mean-square strategy that accounts for communication delays in the inter- and intra-cluster information exchanges. We conduct detailed performance analysis and establish convergence criteria, both in mean and mean-square senses, to investigate the effect of inter-node communication delays on the stability and estimation performance. In particular, we derive convergence conditions and closed-form expressions for transient and steady-state mean square deviation. The concepts are verified using simulation examples, which show a precise match between theoretical and experimental steady-state MSD results.**

***Index Terms*—Distributed estimation, adaptive networks, clustered multitask networks, inter-node communication delays.**

I. Introduction

Initial work on adaptive networks mainly considers a single-task estimation problem where nodes collaboratively estimate a common parameter vector [1]–[4]. In contrast to these single-task networks, multitask adaptive networks consider the problem of estimating multiple parameter vectors based on the information available at different agents [5]. For example, in distributed active noise control, node-specific cooperative spectrum sensing, and node-specific speech enhancement, multiple parameter vectors need to be estimated jointly in a cooperative manner [6]. In clustered multitask networks, groups of nodes form clusters that estimate different parameter vectors. However, if parameter vectors in neighboring clusters are related, the local estimation task's performance can be improved through collaboration across clusters [7].

In this regard, a least mean square (LMS) based multitask diffusion strategy has been presented in [8]. The clustered multitask diffusion LMS strategy in [8] was later extended to asynchronous networks [9], which experience several uncertainties in network links such as topology changes, random link failures, and agents turning on and off to reduce energy consumption. Robust learning approaches under these scenarios have been proposed in [10]. Separately, using the robustness of the affine projection algorithm (APA) against colored input, APA based multitask diffusion schemes were proposed in [11]–[13]. These above single and multitask distributed strategies assume no communications delays during information exchange. However, due to congestion or communication constraints in the network, neighbor messages may arrive at the agent with different time delays. Also, the physical communication medium between nodes may incur delays. For example, underwater acoustic networks encounter inter-node communication delays due to low sound speed. One way to deal with this situation is synchronous diffusion, i.e., all nodes in the network wait for the most prolonged message to complete one cycle of the adaptation and combination process. However, this method slows down the speed of the estimation process. Therefore, it is necessary to study the behavior of distributed strategies under communication delays. Several interesting recent developments on consensus and diffusion strategies in the presence of communication delays have been made [14]–[17]; however, all these strategies deal with single-task estimation.

The clustered multitask diffusion strategies allow the nodes to cooperate on two levels, inter- and intra-cluster levels. Several studies deal with distributed multitask estimation in the presence of random link failures and changing topology [9]. However, the issue of inter-node communication delays in multitask networks has yet to be examined.

In this manuscript, we carry out a detailed analysis of clustered multitask diffusion LMS (CMDLMS) in the presence of inter-node communication delays and provide conditions for its convergence both in the mean and mean-square senses. One of the main findings of this work is that inter-node communication delays do not affect the convergence conditions of clustered multitask diffusion strategies. Furthermore, the simulations indicate that operating on the most recent data exchanges, although subjected to long delays, can avoid a significant slowdown in convergence compared to the synchronous CMDLMS.

II. Clustered multitask diffusion LMS with inter-node communication delays

Consider a sensor network with K number of agents that is modeled as an undirected clustered graph $\mathcal{G} = \{\mathcal{N}, \mathcal{Q}, \mathcal{E}, \}$, where $\mathcal{N}$ is the agent set, $\mathcal{Q} = \{1, 2, \ldots, Q\}$ is the cluster set, and $\mathcal{E}$ is the edge set that represents the bidirectional links between the agents, i.e., $(k, l) \in \mathcal{E}$ if nodes k and l are connected. Furthermore, the set $\mathcal{N}_k$ denotes the neighborhood of node k including itself with cardinality $|\mathcal{N}_k|$. At each time instant n, every node k has access to the input $u_{k,n}$ and the observable output $d_{k,n}$ that are related via a linear model [1]:

$$d_{k,n} = \mathbf{u}_{k,n}^{\mathrm{T}} \mathbf{w}_k^* + \upsilon_{k,n}, \tag{1}$$

where $\mathbf{w}_k^*$ is the $L \times 1$ optimal parameter vector (also termed as task) to be estimated at node k, $\mathbf{u}_{k,n} =$

This work was supported by the Research Council of Norway.

Vinay Chakravarthi Gogineni is with the SimulaMet, Simula Research Laboratory, Norway (e-mail: vcgogineni@simula.no).

Sayed Pouria Talebi and Stefan Werner are with the Norwegian University of Science and Technology, Norway (e-mails:{pouria, stefan.werner}@ntnu.no).

$[u_{k,n}, u_{k,n-1}, \cdots, u_{k,n-L+1}]^T$ is the input regressor at node k and $\upsilon_{k,n}$ is a zero-mean observation noise sequence and variance $\sigma^2_{\upsilon,k}$. In clustered multitask networks, the nodes that are grouped in the same cluster C_q, $q \in \mathcal{Q}$, estimate the same $L \times 1$ coefficient vector $\mathbf{w}^*_{C_q}$, implying $\mathbf{w}^*_k = \mathbf{w}^*_{C_q}$ for $k \in C_q$. Furthermore, the neighboring clusters carry out different but related estimation tasks, implying $\mathbf{w}^\star_{C_q} \sim \mathbf{w}^\star_{C_r}$ if clusters C_q and C_r are connected.

Since similarities exist among the inter-cluster tasks, inter-cluster cooperation, taking place through a suitable regularizer, can improve the overall estimation performance [7], [11]. By using a squared Euclidean-distance regularizer in the cost function, a clustered multitask diffusion LMS (CMDLMS) strategy was proposed in [8]. The adapt-then-combine (ATC) based CMDLMS strategy is given by [8]:

Adaptation:

$$\boldsymbol{\psi}_{k,n} = \mathbf{w}_{k,n-1} + \mu\, \mathbf{u}_{k,n}(d_{k,n} - \mathbf{u}^{\mathrm{T}}_{k,n}\mathbf{w}_{k,n-1}), \tag{2a}$$

Inter-cluster cooperation:

$$\boldsymbol{\psi}'_{k,n} = \boldsymbol{\psi}_{k,n} + \mu\, \eta \sum_{l \in \mathcal{N}_k \setminus C(k)} \rho_{kl}\, (\mathbf{w}_{l,n-1} - \mathbf{w}_{k,n-1}), \tag{2b}$$

Intra-cluster cooperation:

$$\mathbf{w}_{k,n} = \sum_{l \in \mathcal{N}_k \cap C(k)} a_{lk}\, \boldsymbol{\psi}'_{l,n}, \tag{2c}$$

where $\mu > 0$ is the adaptation step size and $\eta > 0$ is the regularization parameter. The symbol $C(k)$ denotes the cluster to which the node k belongs and $\setminus$ represents the set difference operator. Furthermore, the regularizer coefficients ρ_{kl} are non-negative and the matrix $\mathbf{P}$ with $[\mathbf{P}]_{k,l} = \rho_{kl}$ is an asymmetric right-stochastic matrix that defines the regularizer strength among the inter-cluster nodes. The combiner coefficients a_{lk} are non-negative and the matrix $\mathbf{A}$ with $[\mathbf{A}]_{l,k} = a_{lk}$ is a left-stochastic matrix that defines the combining weights of intra-cluster nodes.

The CMDLMS strategy (2) assumes the network to be time-synchronized and that all the intermediate neighbor estimates arrive to the node k before the next iteration. In particular, estimates from inter-cluster neighbors, i.e., $\mathbf{w}_{l,n-1}$ for all $l \in (\mathcal{N}_k \setminus C(k))$, and estimates from intra-cluster neighbors, i.e., $\boldsymbol{\psi}_{l,n}$ for all $l \in (\mathcal{N}_k \cap C(k))$, arrive at node k before the next iteration. However, due to congestion, or, communication constraints in the network, these estimates may arrive after large delays. To capture these practical operating conditions, in this work, we consider the scenario where significant delays may occur during the inter-node communication. This implies that each node k has access to both delayed inter-cluster estimates, i.e., $\mathbf{w}_{l,n-1-\tau_{kl}}$ and delayed intra-cluster estimates, i.e., $\boldsymbol{\psi}_{l,n-\tau_{lk}}$, with $\tau_{kl} = \tau_{lk} \geq 0$ denoting the communication delay from node k to l. After considering the delayed estimates, the ATC based CMDLMS [8] becomes

Adaptation:

$$\boldsymbol{\psi}_{k,n} = \mathbf{w}_{k,n-1} + \mu\, \mathbf{u}_{k,n}(d_{k,n} - \mathbf{u}^{\mathrm{T}}_{k,n}\mathbf{w}_{k,n-1}), \tag{3a}$$

Inter-cluster cooperation:

$$\boldsymbol{\psi}'_{k,n} = \boldsymbol{\psi}_{k,n} + \mu\, \eta \sum_{l \in \mathcal{N}_k \setminus C(k)} \rho_{kl}\, (\mathbf{w}_{l,n-1-\tau_{kl}} - \mathbf{w}_{k,n-1}), \tag{3b}$$

Intra-cluster cooperation:

$$\mathbf{w}_{k,n} = \sum_{l \in \mathcal{N}_k \cap C(k)} a_{lk}\, \boldsymbol{\psi}'_{l,n-\tau_{lk}}. \tag{3c}$$

From (3), notice that at time instant n, each node k uses the delayed estimates from neighbours, i.e., $\mathbf{w}_{l,n-1-\tau_{kl}}$ and $\boldsymbol{\psi}_{l,n-\tau_{lk}}$.

III. Convergence Analysis

A. Error Recursion

Denoting the weight error vectors at node k and time index n as $\widetilde{\boldsymbol{\psi}}_{k,n} = \mathbf{w}^*_k - \boldsymbol{\psi}_{k,n}$ and $\widetilde{\mathbf{w}}_{k,n} = \mathbf{w}^*_k - \mathbf{w}_{k,n}$, the network-level weight error vectors are defined as $\widetilde{\boldsymbol{\psi}}_n = \mathrm{col}\{\widetilde{\boldsymbol{\psi}}_{1,n}, \widetilde{\boldsymbol{\psi}}_{2,n}, \ldots, \widetilde{\boldsymbol{\psi}}_{K,n}\}$ and $\widetilde{\mathbf{w}}_n = \mathrm{col}\{\widetilde{\mathbf{w}}_{1,n}, \widetilde{\mathbf{w}}_{2,n}, \ldots, \widetilde{\mathbf{w}}_{K,n}\}$. Then, we can introduce the following network-level extended error vectors of size $2LKT \times 1$ with $T = \tau + 1$, where $\tau = \max_{k,l \in \mathcal{N}} \tau_{kl}$:

$$\begin{aligned}
\widetilde{\boldsymbol{\psi}}^e_n &\triangleq \mathrm{col}\{\widetilde{\boldsymbol{\psi}}_n, \widetilde{\boldsymbol{\psi}}_{n-1} \ldots, \widetilde{\boldsymbol{\psi}}_{n-\tau}, \widetilde{\mathbf{w}}_{n-1}, \widetilde{\mathbf{w}}_{n-2}, \ldots, \widetilde{\mathbf{w}}_{n-1-\tau}\},\\
\widetilde{\mathbf{w}}^e_n &\triangleq \mathrm{col}\{\widetilde{\mathbf{w}}_n, \widetilde{\mathbf{w}}_{n-1}, \ldots, \widetilde{\mathbf{w}}_{n-\tau}, \widetilde{\boldsymbol{\psi}}_n, \widetilde{\boldsymbol{\psi}}_{n-1}, \ldots, \widetilde{\boldsymbol{\psi}}_{n-\tau}\},\\
\mathbf{w}^*_e &= \mathrm{col}\{\mathbf{w}^*_1, \mathbf{w}^*_2, \ldots, \mathbf{w}^*_K, \mathbf{0}_{LK(2T-1)\times 1}\}.
\end{aligned} \tag{4}$$

Replacing (3) into (4), the recursion for the network-level extended error vector $\widetilde{\mathbf{w}}^e_n$ is obtained as

$$\widetilde{\mathbf{w}}^e_n = \boldsymbol{\mathcal{B}}_n \widetilde{\mathbf{w}}^e_{n-1} - \mu \boldsymbol{\mathcal{A}}^e \mathbf{s}^e_n + \mathbf{r}^e, \tag{5}$$

with

$$\boldsymbol{\mathcal{B}}_n = \boldsymbol{\mathcal{A}}^e\big(\mathbf{I}^e - \mu \boldsymbol{\mathcal{U}}^e_n - \mu\eta \boldsymbol{\mathcal{Q}}^e\big), \tag{6}$$

$$\mathbf{r}^e = \mu\eta \boldsymbol{\mathcal{A}}^e \boldsymbol{\mathcal{Q}}^e \mathbf{w}^*_e, \tag{7}$$

where

$$\begin{aligned}
\mathbf{I}^e &= \begin{bmatrix} \mathbf{I}_{LK} & \mathbf{0}_{LK\times LK(2T-1)} & \\ \mathbf{0}_{LK\tau\times LKT} & \mathbf{I}_{LK\tau} & \mathbf{0}_{LK\tau\times LK} \\ \mathbf{I}_{LKT} & \mathbf{0}_{LKT\times LKT} & \end{bmatrix},\\
\boldsymbol{\mathcal{U}}^e_n &= \begin{bmatrix} \boldsymbol{\mathcal{U}}_n & \mathbf{0}_{LK\times LK(2T-1)} \\ \mathbf{0}_{LK(2T-1)\times LKT} & \mathbf{0}_{LK(2T-1)\times LKT} \end{bmatrix},
\end{aligned} \tag{8}$$

$$\boldsymbol{\mathcal{A}}^e = \begin{bmatrix} \boldsymbol{\mathcal{A}}^{\mathrm{T}}_0 \quad \boldsymbol{\mathcal{A}}^{\mathrm{T}}_1 \quad \ldots \quad \boldsymbol{\mathcal{A}}^{\mathrm{T}}_\tau & \mathbf{0}_{LK\times LKT} & \\ \mathbf{0}_{LK\tau\times LKT} & \mathbf{I}_{LK\tau} & \mathbf{0}_{LK\tau\times LK} \\ \mathbf{I}_{LKT} & \mathbf{0}_{LKT\times LKT} & \end{bmatrix}, \tag{9}$$

$$\begin{aligned}
\boldsymbol{\mathcal{Q}}^e &= \mathbf{I}^{e'} - \boldsymbol{\mathcal{P}}^e \\
&= \mathbf{I}^{e'} - \begin{bmatrix} \boldsymbol{\mathcal{P}}_0 \quad \boldsymbol{\mathcal{P}}_1 \quad \ldots \quad \boldsymbol{\mathcal{P}}_\tau & \mathbf{0}_{LK\times LKT} \\ \mathbf{0}_{LK(2T-1)\times LKT} & \mathbf{0}_{LK(2T-1)\times LKT} \end{bmatrix},
\end{aligned} \tag{10}$$

$$\mathbf{s}^e_n = \mathrm{col}\{\mathbf{u}_{1,n}\upsilon_{1,n}, \mathbf{u}_{2,n}\upsilon_{2,n}, \ldots, \mathbf{u}_{K,n}\upsilon_{K,n}, \mathbf{0}_{LK(2T-1)\times 1}\}, \tag{11}$$

with

$$\boldsymbol{\mathcal{U}}_n = \mathrm{blockdiag}\{\mathbf{u}_{1,n}\mathbf{u}^{\mathrm{T}}_{1,n}, \ldots, \mathbf{u}_{K,n}\mathbf{u}^{\mathrm{T}}_{K,n}\}$$

while $\boldsymbol{\mathcal{A}}_t = \mathbf{A}_t \otimes \mathbf{I}_L$ for $t = 0, 1, \ldots, \tau$, so that $[\mathbf{A}_t]_{l,k} = [\mathbf{A}]_{l,k}$ if $\tau_{l,k} = t$, otherwise $\mathbf{A}_t = \mathbf{0}$. Similarly, $\boldsymbol{\mathcal{P}}_t = \mathbf{P}_t \otimes \mathbf{I}_L$ for $t = 0, 1, \ldots, \tau$, where $[\mathbf{P}_t]_{k,l} = [\mathbf{P}]_{k,l}$ if $\tau_{k,l} = t$, otherwise $\mathbf{P}_t = \mathbf{0}$. The symbol $\otimes$ denotes the Kronecker product. Finally, the matrix $\mathbf{I}^{e'}$ is given by

$$\mathbf{I}^{e'} = \begin{bmatrix} \mathbf{I}_{LK} & \mathbf{0}_{LK\times LK(2T-1)} \\ \mathbf{0}_{LK(2T-1)\times LKT} & \mathbf{0}_{LK(2T-1)\times LKT} \end{bmatrix}. \quad (12)$$

In the following, we study the convergence behavior of the CMDLMS with inter-node communication delays governed by (5). In this regard, the following assumptions are made:

A1: Given a node $k \in \mathcal{N}$, the input regressor $\mathbf{u}_{k,n}$ is drawn from a wide-sense stationary multivariate random sequence with correlation matrix $\mathbf{R}_{u,k} = \mathrm{E}[\mathbf{u}_{k,n}\mathbf{u}_{k,n}^{\mathrm{T}}]$. In addition, the input regressors $\mathbf{u}_{k,n}$ and $\mathbf{u}_{l,m}$ are taken to be independent, for all $k \neq l$ and $m \neq n$.

A2: The observation noise $\upsilon_{k,n}$ is a zero-mean Gaussian random sequence with variance $\sigma^2_{\upsilon,k}$. In addition, $\upsilon_{k,n}$ is taken to be independent of any other random signals in the model.

A3: The step size μ is sufficiently small so that the terms involving higher-order powers of μ can be neglected.

B. First-order Convergence Analysis

Theorem 1. Let **A1**-**A2** hold. Then, the CMDLMS with inter-node communication delays converges in mean provided

$$0 < \mu < \frac{2}{\max\limits_{\forall k\in\mathcal{N}} \big\{ \max\limits_{\forall i}\{\lambda_i(\ \mathbf{R}_{u,k})\}\big\} + 2\ \eta}, \quad (13)$$

where $\lambda_i(\cdot)$ denotes the ith eigenvalue of its argument matrix.

Proof. Taking the statistical expectations $\mathrm{E}[\cdot]$ on both sides of (5) yields

$$E[\widetilde{\mathbf{w}}^e_n] \ = \ \boldsymbol{\mathcal{B}}\, E[\widetilde{\mathbf{w}}^e_{n-1}] + \mathbf{r}^e, \quad (14)$$

where $\boldsymbol{\mathcal{B}} = \mathrm{E}[\boldsymbol{\mathcal{B}}_n] = \boldsymbol{\mathcal{A}}^e\ (\mathbf{I}^e - \mu\ \mathbf{R}^e\ - \mu\ \eta\ \boldsymbol{\mathcal{Q}}^e)$ with $\mathbf{R}^e = \mathrm{E}[\boldsymbol{\mathcal{U}}^e_n] = \mathrm{blockdiag}\{\mathrm{E}[\boldsymbol{\mathcal{U}}_n], \ldots, \mathbf{0}\} = \mathrm{blockdiag}\{\mathbf{R}, \ldots, \mathbf{0}\}$ and $\mathbf{r}^e = \mu\eta\boldsymbol{\mathcal{A}}^e\boldsymbol{\mathcal{Q}}^e\mathbf{w}^*_e$.

From (14), it can be induced that $\lim\limits_{n\to\infty} \mathrm{E}[\widetilde{\mathbf{w}}^e_n]$ attains a finite value if and only if $\|\boldsymbol{\mathcal{B}}\| < 1$, where $\|\cdot\|$ denotes any matrix norm. To derive a convergence condition, we consider the block maximum norm []of the matrix $\boldsymbol{\mathcal{B}}$ (i.e., $\|\boldsymbol{\mathcal{B}}\|_{\mathrm{b},\infty}$). From the properties of the block maximum norm [1], we have

$$\begin{aligned} \|\boldsymbol{\mathcal{B}}\|_{\mathrm{b},\infty} &= \|\boldsymbol{\mathcal{A}}^e\ (\mathbf{I}^e - \mu\ \mathbf{R}^e\ - \mu\ \eta\ \boldsymbol{\mathcal{Q}}^e)\|_{\mathrm{b},\infty} \\ &\leq \|\boldsymbol{\mathcal{A}}^e\|_{b,\infty} \|\mathbf{I}^e - \mu\ \mathbf{R}^e\ - \mu\ \eta\ \boldsymbol{\mathcal{Q}}^e\|_{\mathrm{b},\infty} \\ &= \|\mathbf{I}^e - \mu\ \mathbf{R}^e\ - \mu\ \eta\ \boldsymbol{\mathcal{Q}}^e\|_{\mathrm{b},\infty}. \end{aligned} \quad (15)$$

From the definition of block maximum norm, it follows that $\|\boldsymbol{\mathcal{A}}^e\|_{b,\infty} = 1$, which was used in (15). Substituting (10) in (15) and using the block maximum norm properties, we have

$$\begin{aligned} \|\boldsymbol{\mathcal{B}}\|_{\mathrm{b},\infty} &\leq \|\mathbf{I}^e - \mu\eta\mathbf{I}^{e'} - \mu\ \mathbf{R}^e\ + \mu\ \eta\ \boldsymbol{\mathcal{P}}^e\|_{\mathrm{b},\infty} \\ &\leq \|\mathbf{I}^e - \mu\eta\mathbf{I}^{e'} - \mu\ \mathbf{R}^e\|_{\mathrm{b},\infty}\ + \mu\ \eta\ \|\boldsymbol{\mathcal{P}}^e\|_{\mathrm{b},\infty} \\ &= \|\mathbf{I}^e - \mu\eta\mathbf{I}^{e'} - \mu\ \mathbf{R}^e\|_{\mathrm{b},\infty}\ + \mu\ \eta. \end{aligned} \quad (16)$$

Similar to (15), using the definition of block maximum norm, it is straightforward to prove that $\|\boldsymbol{\mathcal{P}}^e\|_{b,\infty} = 1$. Therefore, a sufficient condition for $\mathrm{E}[\widetilde{\mathbf{w}}^e_n]$ to converge in mean is $\rho\big(\mathbf{I}^e - \mu\eta\mathbf{I}^{e'} - \mu\ \mathbf{R}^e\big) + \mu\eta < 1$, or, equivalently, $\rho\big((1-\mu\eta)\mathbf{I}_{LK} - \mu\ \mathbf{R}\big) + \mu\eta < 1$; which leads to $-1+\mu\eta < 1-\mu\lambda_i(\mathbf{R}) < 1-\mu$ for $i = 1, 2, \ldots, LK$. Thus, a sufficient condition for mean convergence is $0 < \mu < \frac{2}{\max\limits_{i=1,\cdots,LK} \lambda_i(\mathbf{R})+2\eta}$, which concludes the proof. $\square$

C. Second-order Convergence Analysis

Defining the weighted squared-norm of $\widetilde{\mathbf{w}}^e_n$ as $\|\widetilde{\mathbf{w}}^e_n\|^2_{\boldsymbol{\Sigma}} = (\widetilde{\mathbf{w}}^e_n)^{\mathrm{T}}\boldsymbol{\Sigma}\widetilde{\mathbf{w}}^e_n$, where $\boldsymbol{\Sigma}$ is an arbitrary positive semi-definite matrix and under the assumptions **A1** $-$ **A2**, from (5), we have

$$\begin{aligned} \mathrm{E}[\|\widetilde{\mathbf{w}}^e_n\|^2_{\boldsymbol{\Sigma}}] =& \mathrm{E}[\|\widetilde{\mathbf{w}}^e_{n-1}\|^2_{\boldsymbol{\Sigma}'}] + \mu^2\mathrm{E}[(\boldsymbol{\mathcal{A}}^e\mathbf{s}^e_n)^{\mathrm{T}}\boldsymbol{\Sigma}\boldsymbol{\mathcal{A}}^e\mathbf{s}^e_n] \\ &+ \|\mathbf{r}^e\|^2_{\boldsymbol{\Sigma}} + \mathrm{E}[(\boldsymbol{\mathcal{B}}_n\widetilde{\mathbf{w}}^e_{n-1})^{\mathrm{T}}\boldsymbol{\Sigma}\mathbf{r}^e] \\ &+ \mathrm{E}[(\mathbf{r}^e)^{\mathrm{T}}\boldsymbol{\Sigma}\ \boldsymbol{\mathcal{B}}_n\ \widetilde{\mathbf{w}}^e_{n-1}]. \end{aligned} \quad (17)$$

The weighted matrix $\boldsymbol{\Sigma}'$ is given by

$$\boldsymbol{\Sigma}' = \mathrm{E}[\boldsymbol{\mathcal{B}}^{\mathrm{T}}_n\ \boldsymbol{\Sigma}\ \boldsymbol{\mathcal{B}}_n]. \quad (18)$$

Using the block Kronecker product $\otimes_b$ and block vectorization operator $\mathrm{bvec}\{\cdot\}$ [18], we can relate the vectors $\boldsymbol{\sigma} = \mathrm{bvec}\{\boldsymbol{\Sigma}\}$ and $\boldsymbol{\sigma}' = \mathrm{bvec}\{\boldsymbol{\Sigma}'\}$ as

$$\boldsymbol{\sigma}' = \boldsymbol{\mathcal{F}}^{\mathrm{T}}\ \boldsymbol{\sigma}, \quad (19)$$

with

$$\boldsymbol{\mathcal{F}}_n = \mathrm{E}[\boldsymbol{\mathcal{B}}_n \otimes_b \boldsymbol{\mathcal{B}}_n] = \big(\boldsymbol{\mathcal{A}}^e \otimes_b \boldsymbol{\mathcal{A}}^e\big)\ \ \boldsymbol{\mathcal{H}}, \quad (20)$$

where

$$\begin{aligned} \boldsymbol{\mathcal{H}} \approx& (\mathbf{I}^e \otimes_b \mathbf{I}^e) - \mu\big(\mathbf{I}^e \otimes_b \mathbf{R}^e\big) - \mu\big(\mathbf{R}^e \otimes_b \mathbf{I}^e\big) \\ &- \mu\eta\big(\mathbf{I}^e \otimes_b \boldsymbol{\mathcal{Q}}^e\big) - \mu\eta\big(\boldsymbol{\mathcal{Q}}^e \otimes_b \mathbf{I}^e\big). \end{aligned} \quad (21)$$

Under **A3**, the terms involving higher-order powers of μ in (21) become sufficiently small to be non-influential. Hence, the analysis will be continued with this approximation.

The second term on the RHS of (17) can be rewritten as $\mathrm{E}\big[(\boldsymbol{\mathcal{A}}^e\mathbf{s}^e_n)^{\mathrm{T}}\boldsymbol{\Sigma}\boldsymbol{\mathcal{A}}^e\mathbf{s}^e_n\big] = \mathrm{E}[\mathrm{Tr}((\mathbf{s}^e_n)^{\mathrm{T}}(\boldsymbol{\mathcal{A}}^e)^{\mathrm{T}}\boldsymbol{\Sigma}\boldsymbol{\mathcal{A}}^e\mathbf{s}^e_n)] = \mathrm{Tr}(\boldsymbol{\mathcal{A}}^e\mathrm{E}[\mathbf{s}^e_n(\mathbf{s}^e_n)^{\mathrm{T}}](\boldsymbol{\mathcal{A}}^e)^{\mathrm{T}}\boldsymbol{\Sigma})$. Using the assumption **A2**, one can then obtain

$$\mathrm{Tr}(\boldsymbol{\mathcal{A}}^e\mathrm{E}[\mathbf{s}^e_n(\mathbf{s}^e_n)^{\mathrm{T}}](\boldsymbol{\mathcal{A}}^e)^{\mathrm{T}}\boldsymbol{\Sigma}) = \boldsymbol{\gamma}^{\mathrm{T}}\boldsymbol{\sigma}, \quad (22)$$

where

$$\boldsymbol{\gamma} = \mathrm{bvec}\{\boldsymbol{\Phi}\} = \mathrm{bvec}\{\boldsymbol{\mathcal{A}}^e\ \boldsymbol{\mathcal{S}}^e(\boldsymbol{\mathcal{A}}^e)^{\mathrm{T}}\}, \quad (23)$$

with $\boldsymbol{\mathcal{S}}^e = \mathrm{E}[\mathbf{s}^e_n(\mathbf{s}^e_n)^{\mathrm{T}}] = \mathrm{blockdiag}\{\boldsymbol{\mathcal{S}}, \mathbf{0}, \ldots, \mathbf{0}\}$. The quantity $\boldsymbol{\mathcal{S}} = \mathrm{blockdiag}\{\sigma^2_{\nu,1}\mathbf{R}_{u,1}, \sigma^2_{\nu,2}\mathbf{R}_{u,2}, \ldots, \sigma^2_{\nu,K}\mathbf{R}_{u,K}\}$.

The last three terms on the RHS of (17) are evaluated in the following. First,

$$\begin{aligned} \|\mathbf{r}^e\|^2_{\boldsymbol{\Sigma}} &= \mu^2\eta^2\ \mathrm{Tr}(\boldsymbol{\mathcal{A}}^e\boldsymbol{\mathcal{Q}}^e\mathbf{w}^\star_e(\mathbf{w}^\star_e)^{\mathrm{T}}(\boldsymbol{\mathcal{Q}}^e)^{\mathrm{T}}(\boldsymbol{\mathcal{A}}^e)^{\mathrm{T}}\boldsymbol{\Sigma}) \\ &= \mu^2\eta^2\boldsymbol{\alpha}^{\mathrm{T}}\boldsymbol{\sigma}, \end{aligned} \quad (24)$$

where

$$\begin{aligned} \boldsymbol{\alpha} = \mathrm{bvec}\{\boldsymbol{\Delta}^{\mathrm{T}}\} =& \mathrm{bvec}\{\boldsymbol{\mathcal{A}}^e\boldsymbol{\mathcal{Q}}^e\mathbf{w}^\star_e(\mathbf{w}^\star_e)^{\mathrm{T}}(\boldsymbol{\mathcal{Q}}^e)^{\mathrm{T}}(\boldsymbol{\mathcal{A}}^e)^{\mathrm{T}}\} \quad (25) \\ =& (\boldsymbol{\mathcal{A}}^e \otimes_b \boldsymbol{\mathcal{A}}^e)(\boldsymbol{\mathcal{Q}}^e \otimes_b \boldsymbol{\mathcal{Q}}^e)\mathrm{bvec}\{\mathbf{w}^\star_e(\mathbf{w}^\star_e)^{\mathrm{T}}\}. \end{aligned}$$

Next, the term $E[(\boldsymbol{\mathcal{B}}_n\widetilde{\mathbf{w}}^e_{n-1})^{\mathrm{T}}\boldsymbol{\Sigma}\mathbf{r}^e]$ is evaluated as

$$\begin{aligned} \mathrm{E}[(\boldsymbol{\mathcal{B}}_n\widetilde{\mathbf{w}}^e_{n-1})^{\mathrm{T}}\boldsymbol{\Sigma}\mathbf{r}^e] &= \mathrm{E}[\mathrm{bvec}\{(\boldsymbol{\mathcal{B}}_n\widetilde{\mathbf{w}}^e_{n-1})^{\mathrm{T}}\boldsymbol{\Sigma}\mathbf{r}^e\}] \\ &= \mathrm{E}\big[\mathbf{r}^e \otimes_b \boldsymbol{\mathcal{B}}_n\widetilde{\mathbf{w}}^e_{n-1}\big]^{\mathrm{T}}\boldsymbol{\sigma} \\ &= \big((\mathbf{r}^e \otimes_b \boldsymbol{\mathcal{B}})(1 \otimes_b \mathrm{E}[\widetilde{\mathbf{w}}^e_{n-1}])\big)^{\mathrm{T}}\boldsymbol{\sigma} \\ &= \mathrm{E}[(\widetilde{\mathbf{w}}^e_{n-1})^{\mathrm{T}}]\boldsymbol{\Theta}^{\mathrm{T}}\boldsymbol{\sigma}, \end{aligned} \quad (26)$$

where $\boldsymbol{\Theta} = \mathbf{r}^e \otimes_b \boldsymbol{\mathcal{B}} = (\boldsymbol{\mathcal{A}}^e \otimes_b \boldsymbol{\mathcal{A}}^e)\big(\mu\eta(\boldsymbol{\mathcal{Q}}^e \mathbf{w}_e^\star \otimes_b \mathbf{I}^e) - \mu^2\eta(\boldsymbol{\mathcal{Q}}^e \mathbf{w}_e^\star \otimes_b \mathbf{R}^e) - \mu^2\eta^2(\boldsymbol{\mathcal{Q}}^e \otimes_b \boldsymbol{\mathcal{Q}}^e)(\mathbf{w}_e^\star \otimes_b \mathbf{I}^e)\big) \approx \mu\eta(\boldsymbol{\mathcal{A}}^e \otimes_b \boldsymbol{\mathcal{A}}^e)(\boldsymbol{\mathcal{Q}}^e \mathbf{w}_e^\star \otimes_b \mathbf{I}^e)$ (i.e., after neglecting the terms having higher order powers of μ).

Finally, the last term, viz., $\mathrm{E}[(\mathbf{r}^e)^{\mathrm{T}} \boldsymbol{\Sigma} \boldsymbol{\mathcal{B}}_n \widetilde{\mathbf{w}}_{n-1}^e]$ is easily seen to be the same as the previous term, $\mathrm{E}[(\boldsymbol{\mathcal{B}}_n \widetilde{\mathbf{w}}_{n-1}^e)^{\mathrm{T}} \boldsymbol{\Sigma} \mathbf{r}^e]$ evaluated in (26).

From (18)–(26), the expression in (17) can be alternatively formulated as

$$\begin{aligned}\mathrm{E}[\|\widetilde{\mathbf{w}}_n^e\|^2_{\mathrm{bvec}^{-1}\{\boldsymbol{\sigma}\}}] = \mathrm{E}[\|\widetilde{\mathbf{w}}_{n-1}^e\|^2_{\mathrm{bvec}^{-1}\{\boldsymbol{\mathcal{F}}^{\mathrm{T}}\boldsymbol{\sigma}\}}] \\ + \mu^2 \boldsymbol{\gamma}^T \boldsymbol{\sigma} + f\big(\mathrm{E}[\widetilde{\mathbf{w}}_{n-1}^e], \boldsymbol{\sigma}\big),\end{aligned} \tag{27}$$

where $f\big(E[\widetilde{\mathbf{w}}_{n-1}^e], \boldsymbol{\sigma}\big) = \mu^2\eta^2 \boldsymbol{\alpha}^{\mathrm{T}} \boldsymbol{\sigma} + 2\mathrm{E}[(\widetilde{\mathbf{w}}_{n-1}^e)^{\mathrm{T}}] \boldsymbol{\Theta}^{\mathrm{T}} \boldsymbol{\sigma}$. Note that the operator $\mathrm{bvec}^{-1}\{\cdot\}$ rearranges the argument vector of size $4L^2K^2T^2 \times 1$ into a $2LKT \times 2LKT$ matrix, i.e., $\boldsymbol{\Sigma} = \mathrm{bvec}^{-1}\{\boldsymbol{\sigma}\}$.

Theorem 2. Let **A1**-**A3** hold and that (27) describes the dynamics of the weight MSD. Then, the CMDLMS with inter-node communication delays exhibits stable MSD performance provided

$$0 < \mu < \frac{1}{\max\limits_{\forall k \in \mathcal{N}} \big\{ \max\limits_{\forall i} \{\lambda_i(\mathbf{R}_{u,k})\}\big\} + 2\,\eta}. \tag{28}$$

Proof. Iterating (27), backwards down to $n = 0$, we get

$$\begin{aligned}&\mathrm{E}[\|\widetilde{\mathbf{w}}_n^e\|^2_{\mathrm{bvec}^{-1}\{\boldsymbol{\sigma}\}}] = \mathrm{E}[\|\widetilde{\mathbf{w}}_{-1}^e\|^2_{\mathrm{bvec}^{-1}\{(\boldsymbol{\mathcal{F}}^{\mathrm{T}})^{n+1}\,\boldsymbol{\sigma}\}}] \\ &+ \mu^2\,\boldsymbol{\gamma}^T \Big(\mathbf{I}_{4L^2K^2T^2} + \sum_{j=1}^{n} (\boldsymbol{\mathcal{F}}^{\mathrm{T}})^j\Big) \boldsymbol{\sigma} + f(\mathrm{E}[\widetilde{\mathbf{w}}_{n-1}^e], \boldsymbol{\sigma}) \\ &+ \sum_{j=1}^{n} f(\mathrm{E}[\widetilde{\mathbf{w}}_{n-1-j}^e], (\boldsymbol{\mathcal{F}}^{\mathrm{T}})^j \boldsymbol{\sigma}),\end{aligned} \tag{29}$$

where $\widetilde{\mathbf{w}}_{-1}^e = \mathbf{w}_e^* - \mathbf{w}_{-1}^e$. Under $\rho\big(\boldsymbol{\mathcal{F}}^{\mathrm{T}}\big) < 1$, in the steady-state, the first and second terms in the RHS of (29) converge to zero and a finite value, respectively. Since $\mathrm{E}[\widetilde{\mathbf{w}}_{n-1}^e)]$ is bounded, the term $\mathbf{f}(\mathrm{E}[\widetilde{\mathbf{w}}_{n-1-j}], (\boldsymbol{\mathcal{F}}^{\mathrm{T}})^j \boldsymbol{\sigma})$ attains a finite value iff the spectral radius of the matrix $\boldsymbol{\mathcal{F}}$, i.e., $\rho\big(\boldsymbol{\mathcal{F}}^{\mathrm{T}}\big)$ is less than one. Therefore, the convergence of $\mathrm{E}[\|\widetilde{\mathbf{w}}_n^e\|^2_{\boldsymbol{\Sigma}}] = E[\|\widetilde{\mathbf{w}}_n^e\|^2_{\mathrm{bvec}^{-1}\{\boldsymbol{\sigma}\}}]$ requires $\rho\big(\boldsymbol{\mathcal{F}}\big) = \rho\big(\boldsymbol{\mathcal{F}}^T\big) < 1$. From the properties of block maximum norm, we can write

$$\rho\big(\boldsymbol{\mathcal{F}}\big) \le \big\|\big(\boldsymbol{\mathcal{A}}^e \otimes_b \boldsymbol{\mathcal{A}}^e\big)\,\boldsymbol{\mathcal{H}}\|_{b,\infty} \le \|\boldsymbol{\mathcal{A}}^e \otimes_b \boldsymbol{\mathcal{A}}^e\|_{b,\infty}\,\|\boldsymbol{\mathcal{H}}\|_{b,\infty}. \tag{30}$$

From the definition of the block maximum norm, it can be easily shown $\|\boldsymbol{\mathcal{A}}^e \otimes_b \boldsymbol{\mathcal{A}}^e\|_{b,\infty} = 1$. Using this result and substituting $\boldsymbol{\mathcal{H}}$ as given in (21) where $\boldsymbol{\mathcal{Q}}^e$ is given by (10), we have

$$\begin{aligned}\rho\big(\boldsymbol{\mathcal{F}}\big) &\le \|\boldsymbol{\mathcal{X}}\|_{b,\infty} + \mu\eta\,\|(\boldsymbol{\mathcal{P}} \otimes \mathbf{I}^e) + (\mathbf{I}^e \otimes \boldsymbol{\mathcal{P}})\|_{b,\infty} \\ &\le \|\boldsymbol{\mathcal{X}}\|_{b,\infty} + 2\mu\eta,\end{aligned} \tag{31}$$

where $\boldsymbol{\mathcal{X}} = (\mathbf{I}^e \otimes_b \mathbf{I}^e) - \mu\eta(\mathbf{I}^e \otimes_b \mathbf{I}^{e'}) - \mu\eta(\mathbf{I}^{e'} \otimes_b \mathbf{I}^e)$ $- \mu(\mathbf{I}^e \otimes_b \mathbf{R}^e) - \mu(\mathbf{R}^e \otimes_b \mathbf{I}^e)$. Similar to the above, from the definition of block maximum norm, it can be easily shown $\|\boldsymbol{\mathcal{P}} \otimes \mathbf{I}^e\|_{b,\infty} = \|\mathbf{I}^e \otimes \boldsymbol{\mathcal{P}}\|_{b,\infty} = 1$. Therefore, it is seen that the convergence of $E[\|\widetilde{\mathbf{w}}_n^e\|^2_{\boldsymbol{\Sigma}}]$ requires $\rho\big(\boldsymbol{\mathcal{X}}\big) + 2\mu\eta < 1$, or, equivalently, $\rho\big((1 - 2\mu\eta)\mathbf{I}_{L^2K^2} - \mu(\mathbf{I}_{LK} \otimes_b \mathbf{R}) - \mu(\mathbf{R} \otimes_b \mathbf{I}_{LK})\big) + 2\mu\eta < 1$; which leads to $-1 + 2\mu\eta < (1 - 2\mu\eta) - \mu(\lambda_i(\mathbf{R}) + \lambda_j(\mathbf{R})) < 1 - 2\mu\eta$, $i, j = 1, 2, \cdots, LK$. Thus, a sufficient condition for convergence is given by $0 < \mu < \frac{2}{4\eta + 2 \max\limits_{i=1,\cdots,LK} \lambda_i(\mathbf{R})}$, which proves (28). $\square$

Remark 1. The convergence conditions of the CMDLMS with inter-node communication delays are the same as that of the conventional CMDLMS [8] (i.e., with ideal communications).

D. Transient and Steady-State Mean Square Deviation

Using (29), $\mathrm{E}[\|\widetilde{\mathbf{w}}_n^e\|^2_{\mathrm{bvec}^{-1}\{\boldsymbol{\sigma}\}}]$ and $\mathrm{E}[\|\widetilde{\mathbf{w}}_{n-1}^e\|^2_{\mathrm{bvec}^{-1}\{\boldsymbol{\sigma}\}}]$ can be related as

$$\begin{aligned}&\mathrm{E}[\|\widetilde{\mathbf{w}}_n^e\|^2_{\mathrm{bvec}^{-1}\{\boldsymbol{\sigma}\}}] = \mathrm{E}[\|\widetilde{\mathbf{w}}_{n-1}^e\|^2_{\mathrm{bvec}^{-1}\{\boldsymbol{\sigma}\}}] + \mu^2 \boldsymbol{\gamma}^{\mathrm{T}} (\boldsymbol{\mathcal{F}}^{\mathrm{T}})^n\,\boldsymbol{\sigma} \\ &+\mathrm{E}\big[\|\widetilde{\mathbf{w}}_{-1}^e\|^2_{\mathrm{bvec}^{-1}\{(\boldsymbol{\mathcal{F}}^{\mathrm{T}} - \mathbf{I}_{4L^2K^2T^2})(\boldsymbol{\mathcal{F}}^{\mathrm{T}})^n \boldsymbol{\sigma}\}}\big] + \mu^2\eta^2 \boldsymbol{\alpha}^{\mathrm{T}} (\boldsymbol{\mathcal{F}}^{\mathrm{T}})^n \boldsymbol{\sigma} \\ &+2\mathrm{E}[(\widetilde{\mathbf{w}}_{n-1}^e)^{\mathrm{T}}] \boldsymbol{\Theta}^{\mathrm{T}} \boldsymbol{\sigma} + 2\mathrm{E}[(\widetilde{\mathbf{w}}_{n-2}^e)^{\mathrm{T}}] \boldsymbol{\Theta}^{\mathrm{T}} (\boldsymbol{\mathcal{F}}^{\mathrm{T}} - \mathbf{I}_{4L^2K^2T^2}) \boldsymbol{\sigma} \\ &+2 \sum_{j=1}^{n-1} \mathrm{E}[(\widetilde{\mathbf{w}}_{n-2-j}^e)^{\mathrm{T}}] \boldsymbol{\Theta}^{\mathrm{T}} (\boldsymbol{\mathcal{F}}^{\mathrm{T}} - \mathbf{I}_{4L^2K^2T^2})\,(\boldsymbol{\mathcal{F}}^{\mathrm{T}})^j\,\boldsymbol{\sigma}.\end{aligned} \tag{32}$$

Let $\boldsymbol{\sigma} = \mathrm{bvec}\{\mathrm{blockdiag}\{\mathbf{I}_{LK}, \mathbf{0}, \ldots, \mathbf{0}\}\}$, approximating $\boldsymbol{\mathcal{F}} \approx \boldsymbol{\mathcal{B}} \otimes_b \boldsymbol{\mathcal{B}}$, the network-level mean square deviation (MSD) at time index n $\zeta_n = \frac{1}{K} E[\|\widetilde{\mathbf{w}}_n\|^2] = \frac{1}{K} E[\|\widetilde{\mathbf{w}}_n^e\|^2_{\mathrm{bvec}^{-1}\{\boldsymbol{\sigma}\}}]$ evolves as

$$\begin{aligned}\zeta_n &= \zeta_{n-1} + \mu^2 \mathrm{Tr}(\boldsymbol{\Phi}(\boldsymbol{\mathcal{B}}^{\mathrm{T}})^n \boldsymbol{\Sigma} \boldsymbol{\mathcal{B}}^n) \\ &+ \mathrm{Tr}(\widetilde{\mathbf{w}}_{-1}^e (\widetilde{\mathbf{w}}_{-1}^e)^{\mathrm{T}} (\boldsymbol{\mathcal{B}}^{\mathrm{T}})^{n+1} \boldsymbol{\Sigma} \boldsymbol{\mathcal{B}}^{n+1}) \\ &- \mathrm{Tr}(\widetilde{\mathbf{w}}_{-1}^e (\widetilde{\mathbf{w}}_{-1}^e)^{\mathrm{T}} (\boldsymbol{\mathcal{B}}^{\mathrm{T}})^n \boldsymbol{\Sigma} \boldsymbol{\mathcal{B}}^n) + \mu^2\eta^2 \mathrm{Tr}(\boldsymbol{\Delta}(\boldsymbol{\mathcal{B}}^{\mathrm{T}})^n \boldsymbol{\Sigma} \boldsymbol{\mathcal{B}}^n) \\ &+ 2\mathrm{Tr}(\mathbf{r}^e E[(\widetilde{\mathbf{w}}_{n-1}^e)^{\mathrm{T}}] \boldsymbol{\mathcal{B}}^{\mathrm{T}} \boldsymbol{\sigma}) + 2\mathrm{Tr}(\mathbf{r}^e E[(\widetilde{\mathbf{w}}_{n-2}^e)^{\mathrm{T}}] (\boldsymbol{\mathcal{B}}^2)^{\mathrm{T}} \boldsymbol{\Sigma} \boldsymbol{\mathcal{B}}) \\ &- 2\mathrm{Tr}(\mathbf{r}^e E[(\widetilde{\mathbf{w}}_{n-2}^e)^{\mathrm{T}}] \boldsymbol{\mathcal{B}}^{\mathrm{T}} \boldsymbol{\sigma}) \\ &+ 2 \sum_{j=1}^{n-1} \mathrm{Tr}(\mathbf{r}^e E[(\widetilde{\mathbf{w}}_{n-2-j}^e)^{\mathrm{T}}] (\boldsymbol{\mathcal{B}}^{j+2})^{\mathrm{T}} \boldsymbol{\Sigma} \boldsymbol{\mathcal{B}}^{j+1}) \\ &+ 2 \sum_{j=1}^{n-1} \mathrm{Tr}(\mathbf{r}^e E[(\widetilde{\mathbf{w}}_{n-2-j}^e)^{\mathrm{T}}] (\boldsymbol{\mathcal{B}}^{j+1})^{\mathrm{T}} \boldsymbol{\Sigma} \boldsymbol{\mathcal{B}}^j).\end{aligned}$$

Under (28), letting $n \to \infty$ on both sides of (27), we obtain

$$\lim_{n \to \infty} \mathrm{E}[\|\widetilde{\mathbf{w}}_n^e\|^2_{\mathrm{bvec}^{-1}\{(\mathbf{I}_{4L^2K^2T^2} - \boldsymbol{\mathcal{F}}^{\mathrm{T}})\boldsymbol{\sigma}\}}] = \mu^2 \boldsymbol{\gamma}^{\mathrm{T}} \boldsymbol{\sigma} + \mathbf{f}\,(\mathrm{E}[\widetilde{\mathbf{w}}_\infty^e], \boldsymbol{\sigma})\,, \tag{33}$$

where $\mathrm{E}[\widetilde{\mathbf{w}}_\infty^e] = \lim\limits_{n \to \infty} \mathrm{E}[\widetilde{\mathbf{w}}_n^e]$. By substituting $\boldsymbol{\sigma} = \frac{1}{K}$ $(\mathbf{I}_{4L^2K^2T^2} - \boldsymbol{\mathcal{F}}_\delta^{\mathrm{T}})^{-1}\,\mathrm{bvec}\{\mathrm{blockdiag}\{\mathbf{I}_{LK}, \mathbf{0}, \ldots, \mathbf{0}\}\}$ in (33), the network-level steady-state MSD can be obtained.

IV. Simulation Results

This section validates the analytical results and studies the impact of inter-node communication delays on the CMDLMS. For this purpose, we consider a clustered multitask network consisting of $N = 30$ nodes that are grouped in to 4 clusters with the topology shown in Fig. 1(a). These clusters aim to estimate their respective 10-tap parameter vectors in a collaborative fashion which are chosen as $\mathbf{w}_{C_q}^* = \mathbf{w}_0 + \delta_{C_q} \mathbf{w}_{C(k)}$ for $q = 1, 2, 3, 4$ with $\delta_{C_1} = 0, \delta_{C_2} = -0.03$, $\delta_{C_3} = 0.05$ and $\delta_{C_4} = -0.05$. The vectors $\mathbf{w}_0$ and $\mathbf{w}_{\mathcal{C}(k)}$ were generated from a zero-mean, unit variance Gaussian distribution. The input signal $u_{k,n}$ and the noise signal $\upsilon_{k,n}$ at each node are white Gaussian with node-dependent variances $\sigma_{u,k}^2$ and $\sigma_{\upsilon,k}^2$, respectively. The coefficients ρ_{kl} and a_{lk} are set similar to [8].

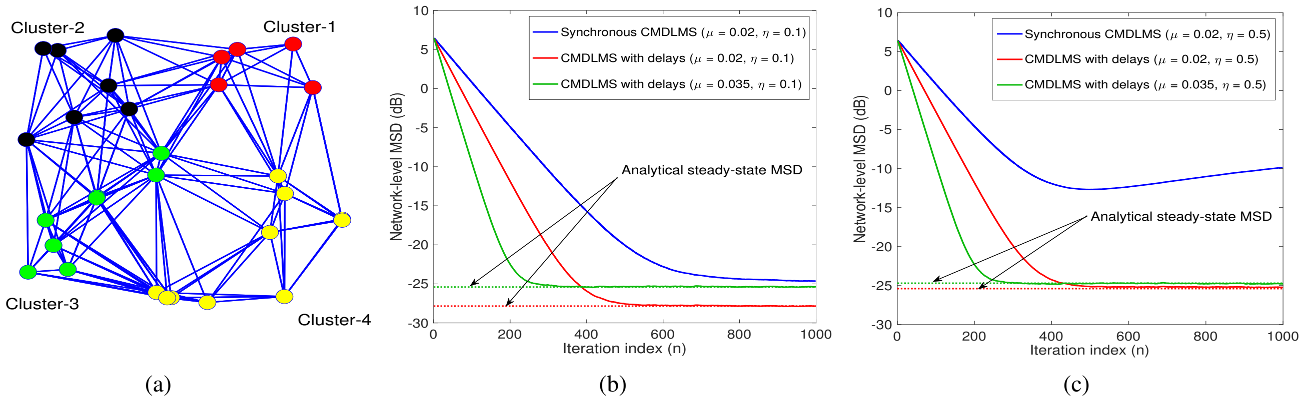

Fig. 1: (a). Network. (b)-(c). MSD comparison.

The inter-node communication delays are proportional to the distances among nodes.

For comparative assessment, the same identification exercise was also carried out by synchronous CMDLMS (in which all nodes wait for the most delayed communication to complete one cycle of adaptation and combination process). Figs. 1(b)-1(c) show the network-level MSD (in dB) against the iteration index n, obtained by averaging over 500 independent experiments. The resulting plots are shown in Figs. 1(b) and 1(c). We see that the CMDLMS with delays exhibited a faster convergence rate and lower steady-state MSD than the synchronous CMDLMS, as it completes the inter-cluster and intra-cluster cooperations with the available information without waiting for the prolonged messages. Moreover, it is well known that the performance of CMDLMS is strongly dependent on the value of η [11], i.e., after a certain value of η the performance of CMDLMS starts deteriorating, which can be observed in the case of synchronous CMDLMS shown in Fig. 1(c). However, in the case of CMDLMS with inter-node communication delays, since the inter-cluster cooperation is running on delayed estimates, the effect of high η values on the network-level MSD is reduced. Hence, CMDLMS with delays shows robustness against the η value as evident in Fig. 1(c). The simulation results precisely match the theoretical steady-state MSD results. Numerous simulations for different step sizes obeying A4, not shown here, corroborate the preciseness of (13),(28), and (32).

V. Conclusions

Distributed adaptive estimation over clustered multitask networks in the presence of inter-node communication delays has been considered. The performance of the clustered multitask diffusion LMS with delayed estimate exchanges was analyzed, and convergence conditions in the mean and mean-square senses were established. Closed-form expressions have been obtained for both network-level transient and steady-state mean square deviation. Simulation results have shown good agreement with theoretical findings.

References

[1] A. H. Sayed, *Adaptation, learning, and optimization over networks*, Boston-Delft: NOW, 2014

[2] A. H. Sayed, "Adaptive networks," *Proc. of the IEEE*, vol. 102, no. 4, pp. 460-497, Apr. 2014.

[3] A. Rastegarnia, "Reduced-communication diffusion RLS for distributed estimation over multi-agent networks," *IEEE Trans. Circuits Syst. II: Express Briefs*, vol. 67, no. 1, pp. 177-181, Jan. 2020.

[4] H. Zayyani, "Robust minimum disturbance diffusion LMS for distributed estimation," *IEEE Trans. Circuits Syst. II: Express Briefs*, doi: 10.1109/TCSII.2020.3004507.

[5] J. Plata-Chaves, N. Bogdanović and K. Berberidis, "Distributed diffusion-based LMS for node-specific adaptive parameter estimation," *IEEE Trans. Signal Process.*, vol. 63, no. 13, pp. 3448-3460, Jul. 2015.

[6] A. Hassani, J. Plata-Chaves, M. H. Bahari, M. Moonen and A. Bertrand, "Multi-task wireless sensor network for joint distributed node-specific signal enhancement, LCMV beamforming and DOA estimation," *IEEE J. Sel. Topics Signal Process.*, vol. 11, no. 3, pp. 518-533, Apr. 2017.

[7] J. Chen, C. Richard and A. H. Sayed, "Diffusion LMS for clustered multitask networks," in *Proc. IEEE Int. Conf. on Acoust., Speech and Signal Process.*, Florence, 2014, pp. 5487-5491,

[8] 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.

[9] 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.

[10] M. Hajiabadi, G. A. Hodtani and H. Khoshbin, "Robust learning over multitask adaptive networks with wireless communication links," *IEEE Trans. Circuits Syst. II: Exp. Briefs*, vol. 66, no. 6, pp. 1083-1087, Jun. 2019.

[11] V. C Gogineni and M. Chakraborty, "Diffusion affine projection algorithm for multitask networks," in *Proc. IEEE Int. Conf. Asia Pacific Signal and Info. Process. Assoc.*, Honolulu, 2018, pp. 201-206.

[12] V. C. Gogineni and M. Chakraborty, "Improving the performance of multitask diffusion APA via controlled inter-cluster cooperation," *IEEE Trans. Circuits Syst. I: Reg. Papers*, vol. 67, no. 3, pp. 903-912, Mar. 2020.

[13] V. C. Gogineni and M. Chakraborty, "Partial diffusion affine projection algorithm over clustered multitask networks," in *Proc. IEEE Int. Symp. on Circuits and Syst. (ISCAS)*, Sapporo, May 2019, pp. 1-5.

[14] K. I. Tsianos and M. G. Rabbat, "Distributed consensus and optimization under communication delays," in *Proc. Annu. Allerton Conf. Commun. Control Comput.*, Monticello, Sep. 2011, pp. 974–982.

[15] T. Wu, K. Yuan, Q. Ling, W. Yin and A. H. Sayed, "Decentralized consensus optimization with asynchrony delays,"*IEEE Trans. Signal Inf. Process. Netw.*, vol. 4, no. 2, pp. 293–307, Jun. 2018.

[16] F. Chen, L. Hu, P. Liu and M. Feng, "A robust diffusion estimation algorithm for asynchronous networks in IoT," *IEEE Internet Things J.*, vol. 7, no. 9, pp. 9103-9115, Sep. 2020.

[17] F. Hua, R. Nassif, C. Richard, H. Wang and A. H. Sayed, "Diffusion LMS with communication delays: Stability and performance analysis,"*IEEE Signal Process. Lett.*, vol. 27, pp. 730-734, Apr. 2020.

[18] R. H.Koning, H. Neudecker and T. Wansbeek, "Block Kronecker products and the vecb operator," *Linear Algebra and its Applications*, vol. 149, pp. 165-184, 1991.