CLOSED-FORM MSE PERFORMANCE OF THE DISTRIBUTED LMS ALGORITHM

Gonzalo Mateos, Ioannis D. Schizas and Georgios B. Giannakis

Dept. of ECE, Univ. of Minnesota, 200 Union Street SE, Minneapolis, MN 55455, USA

ABSTRACT

Mean-square error (MSE) performance analysis is conducted for a novel *distributed* least-mean square (D-LMS) algorithm, which is based on consensus, in-network, adaptive estimation using wireless sensor networks (WSNs). For sensor observations that are linearly related to the time-invariant parameter of interest and independent Gaussian data, exact closed-form expressions are derived for the global and sensor-level MSE evolution and steady-state limiting values. Tracking performance is also investigated when the true parameter adheres to a random-walk model. Remarkably, for small step-sizes the results accurately extend to the pragmatic setup whereby sensors acquire temporally-correlated (non-)Gaussian data.

Index Terms— Wireless sensor networks, distributed estimation, LMS algorithm, performance analysis.

1. INTRODUCTION

It has been recognized that sensors comprising WSNs deployed to perform collaborative estimation tasks, should be empowered with signal processing tools that enable estimation in constantly changing environments, without having available a complete statistical description of the underlying processes of interest. Emergent WSN-based applications include distributed localization, power spectrum estimation, target tracking and have motivated the development of *distributed adaptive* estimation schemes; for early work see e.g., [1] and references therein. Acknowledging such a challenge, a consensus-based distributed (D-)LMS algorithm was developed in [2] for in-network adaptive processing using WSNs with noisy links.

The present paper complements [2] by conducting a detailed MSE performance analysis. For a time-invariant parameter and sensor observations that are linearly related to it, we rely on the simplifying (though generally unrealistic) white Gaussian setting assumptions [3, pg. 246], [4, pg. 110] to derive closed-form expressions for the global and sensor-level MSE evolution and steady-state values. Mean and MSE stability are also assessed. The importance of such results is threefold: (i) an exact MSE characterization is provided for D-LMS, a complex time-varying stochastic dynamical system; (ii) similar existing results for the diffusion LMS [1] lay a common ground for fair comparisons; and (iii) for small step-sizes the conclusions extend to temporally-correlated non-Gaussian data.

The second major contribution of this paper pertains to performance analysis of tracking systems, which to the best of our knowledge is novel in the distributed adaptive estimation literature. Due to its simplicity and widespread use in the analysis of classical LMS [3, pg. 270], [4, pg. 121], we adopt a random-walk model to describe the fluctuations of the time-varying parameter. Though very simple and somehow contrived due to its increasing variance, such model proves sufficient to provide valuable insight about the tracking behavior of D-LMS.

Notation: Operators $\otimes$, $\circ$, $(.)^T$, $(.)^\dagger$, $\lambda_{\max}(.)$, $\text{tr}(.)$, $\text{diag}(.)$, $\text{bdiag}(.)$, $E[.]$ will denote Kronecker product, Hadamard product, transposition, matrix pseudo-inverse, spectral radius, matrix trace, diagonal matrix, block diagonal matrix and expectation, respectively. For both vector and matrices, $\|.\|$ will stand for the $2-$norm and $|.|$ for the cardinality of a set. The $n \times n$ identity matrix will be represented by $\mathbf{I}_n$, while $\mathbf{1}_n$ will denote the $n \times 1$ vector of all ones and $\mathbf{b}_{n,i}$ will stand for the i-th vector in the canonical basis for $\mathbb{R}^n$.

2. PROBLEM STATEMENT AND D-LMS ALGORITHM

Consider a WSN with sensors $\{1, \ldots, J\} := \mathcal{J}$. Only single-hop communications are allowed, i.e., sensor j communicates only with its neighbors in $\mathcal{N}_j \subseteq \mathcal{J}$. Assuming that inter-sensor links are symmetric, the WSN is modeled as an undirected connected graph with adjacency matrix $\mathbf{E}$. Different from [1], the present network model accounts explicitly for non-ideal sensor-to-sensor links, through a zero-mean additive noise vector $\boldsymbol{\eta}_j^i(t)$ with covariance matrix $\mathbf{R}_{\eta_{j,i}} := E[\boldsymbol{\eta}_j^i(t)\boldsymbol{\eta}_j^i(t)^T]$ corrupting signals received at sensor j from sensor i at discrete-time instant t. The noise vectors $\{\boldsymbol{\eta}_j^i(t)\}_{j \in \mathcal{J}}^{i \in \mathcal{N}_j}$ are assumed temporally and spatially uncorrelated.

The WSN is deployed to estimate a signal vector $\mathbf{s}_0(t) \in \mathbb{R}^{p \times 1}$ in a distributed fashion. Per time instant $t = 0, 1, 2, \ldots$, each sensor has available a regression vector $\mathbf{h}_j(t) \in \mathbb{R}^{p \times 1}$ and a scalar observation $x_j(t)$, both assumed zero-mean without loss of generality. A similar data setting was considered in [1]. The global LMS estimator of interest can be written as [3, p. 62], [1], [2]

$$\hat{\mathbf{s}}(t) = \arg \min_{\mathbf{s}} \sum_{j=1}^{J} E\left[(x_j(t) - \mathbf{h}_j^T(t)\mathbf{s})^2\right]. \tag{1}$$

Next we describe an application setup for distributed linear regression, which naturally gives rise to the aforementioned data setting.

2.1. Distributed Power Spectrum Estimation

Consider an ad hoc WSN deployed e.g., for collaborative habitat monitoring, whereby sensors observe a narrowband source to determine its spectral peaks. Such information enables the WSN to disclose hidden periodicities due to a physical phenomenon controlled by e.g., a natural heat source. Denote the source of interest by $\theta(t)$, which can be modeled as an autoregressive (AR) process [5, pg. 106]

$$\theta(t) = -\sum_{\tau=1}^{p} \alpha_\tau \theta(t-\tau) + w(t) \tag{2}$$

Work in this paper was supported by the USDoD ARO Grant No. W911NF-05-1-0283; and also through collaborative participation in the C&N Consortium sponsored by the U. S. ARL under the CTA Program, Cooperative Agreement DAAD19-01-2-0011. The U. S. Government is authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation thereon.

where p is the order of the AR process, while $\{\alpha_\tau\}$ are the AR coefficients and $w(t)$ denotes white noise. The source propagates to sensor j via a multi-path channel modeled as an FIR filter $C_j(z) = \sum_{l=0}^{L_j-1} c_{j,l} z^{-l}$, of *unknown* order L_j and tap coefficients $\{c_{j,l}\}$. In the presence of additive sensing noise $\bar{\epsilon}_j(t)$ the observation at sensor j is thus $x_j(t) = \sum_{l=0}^{L_j-1} c_{j,l}\theta(t-l) + \bar{\epsilon}_j(t)$. Since $x_j(t)$ is an ARMA process, it can be written as [5]

$$x_j(t) = -\sum_{\tau=1}^{p} \alpha_\tau x_j(t-\tau) + \sum_{\tau'=1}^{m} \beta_{\tau'} \bar{\xi}_j(t-\tau'), \quad j \in \mathcal{J} \quad (3)$$

where the moving average (MA) coefficients $\{\beta_{\tau'}\}$ and the variance of the white noise process $\bar{\xi}_j(t)$ depend on $\{c_{j,l}\}$, $\{\alpha_\tau\}$ and the variance of the noise terms $w(t)$ and $\bar{\epsilon}_j(t)$. For the purpose of determining spectral peaks, the MA term in (3) can be treated as observation noise, i.e., $\epsilon_j(t) := \sum_{\tau'=1}^{m} \beta_{\tau'} \bar{\xi}_j(t-\tau')$. This is important since sensors do not have to know the source-sensor channel coefficients as well as the noise variances. The spectral content of the source can be estimated provided sensors estimate the coefficients $\{\alpha_\tau\}$, so let $\mathbf{s}_0 := [\alpha_1 \ldots \alpha_p]^T$. From (3) the regressor vectors are given as $\mathbf{h}_j(t) = [-x_j(t-1) \ldots - x_j(t-p)]^T$, directly from the sensor data $\{x_j(t)\}$ without the need of training/estimation.

Remark 1 The source-sensor channels may introduce deep fades at the frequencies occupied by the source. Thus, having each sensor operating on its own may lead to faulty assessments. The necessary spatial diversity to effect improved spectral estimates, can only be achieved via sensor collaboration as in D-LMS described next.

2.2. The D-LMS Algorithm

To distribute the cost function in (1) two steps are needed: (i) replace the coupling variable $\mathbf{s}$ with auxiliary local variables $\{\mathbf{s}_j\}_{j=1}^J$ that represent candidate estimates of $\mathbf{s}$ per sensor; and (ii) add the constraints $\mathbf{s}_j = \bar{\mathbf{s}}_b$, $b \in \mathcal{B}, j \in \mathcal{N}_b$ where $\mathcal{B} \subseteq \mathcal{J}$ is the *bridge* sensor subset [2]. For future reference, let $\mathcal{B}_j := \mathcal{N}_j \cap \mathcal{B}$ denote the set of bridge neighbors of sensor j. In addition to $\mathbf{s}_b$, bridge sensor $b \in \mathcal{B}$ *also* maintains the local vector $\bar{\mathbf{s}}_b$ utilized to impose consensus among all $\{\mathbf{s}_j\}_{j=1}^J$. Associating Lagrange multipliers $\{\mathbf{v}_j^b\}_{j\in\mathcal{J}}^{b\in\mathcal{B}_j}$ to the aforementioned constraints, the resulting equivalent *constrained* optimization is solved in a distributed fashion by resorting to the alternating-direction method of multipliers [6, p. 253] and a stochastic approximation iteration to handle the unavailability/variation of statistical information. The resultant D-LMS algorithm comprises the following simple recursions updated at all sensors $j \in \mathcal{J}$ and $b \in \mathcal{B}$

$$\mathbf{v}_j^b(t) = \mathbf{v}_j^b(t-1) + c\left(\mathbf{s}_j(t) - (\bar{\mathbf{s}}_b(t) + \boldsymbol{\eta}_j^b(t))\right) \quad (4)$$

$$\mathbf{s}_j(t+1) = \mathbf{s}_j(t) + \mu[2\mathbf{h}_j(t+1)e_j(t+1) - \sum_{b\in\mathcal{B}_j}\left(\mathbf{v}_j^b(t) + c\left(\mathbf{s}_j(t) - (\bar{\mathbf{s}}_b(t) + \boldsymbol{\eta}_j^b(t))\right)\right)] \quad (5)$$

$$\bar{\mathbf{s}}_b(t+1) = \sum_{j\in\mathcal{N}_b} \frac{c^{-1}\mathbf{v}_j^b(t) + \mathbf{s}_j(t+1) + \bar{\boldsymbol{\eta}}_b^j(t+1)}{|\mathcal{N}_b|} \quad (6)$$

where $\mu > 0$ is a constant step-size, $c > 0$ is a penalty coefficient and $e_j(t) := x_j(t) - \mathbf{h}_j^T(t)\mathbf{s}_j(t-1)$ is the local *a priori* error. The overall operation of the algorithm can be described as follows. At time instant t, sensor j receives the (noise corrupted) consensus variables $\bar{\mathbf{s}}_b(t) + \boldsymbol{\eta}_j^b(t)$ from its bridge neighbors $b \in \mathcal{B}_j$. Utilizing (4), it is able to update its Lagrange multipliers $\{\mathbf{v}_j^b(t)\}_{b\in\mathcal{B}_j}$ which are then used to compute $\mathbf{s}_j(t+1)$ via (5). Finally, sensor j transmits the quantity $c^{-1}\mathbf{v}_j^b(t) + \mathbf{s}_j(t+1)$ to all bridge sensors in its neighborhood $\mathcal{B}_j$. Consequently, each bridge sensor $b \in \mathcal{B}$ acquires the vectors $\{c^{-1}\mathbf{v}_j^b(t) + \mathbf{s}_j(t+1) + \bar{\boldsymbol{\eta}}_b^j(t+1)\}_{j\in\mathcal{N}_b}$ whose average is computed using (6) to yield $\bar{\mathbf{s}}_b(t+1)$, thus completing the t-th iteration. For a detailed description of the algorithm's communication and computational cost analyses as well as and comparisons with the diffusion LMS [1], the reader is referred to [2].

3. ANALYSIS PRELIMINARIES

Our approach to performance analysis relies on a compact *error-form* representation of D-LMS as a linear time-varying stochastic difference-equation [2, Lemmata 2 and 3].

3.1. Error-form D-LMS

In this subsection, we start from the D-LMS recursions in (4)-(6) and characterize the evolution of the local estimation errors $\{\mathbf{y}_{1,j}(t) := \mathbf{s}_j(t) - \mathbf{s}_0(t)\}_{j=1}^J$ and local sum of multipliers $\{\mathbf{y}_{2,j}(t) := \sum_{b\in\mathcal{B}_j}\mathbf{v}_j^b(t-1)\}_{j=1}^J$. A convenient global state capturing the spatio-temporal dynamics of D-LMS can be defined as $\mathbf{y}(t) := [\mathbf{y}_1^T(t)\ \mathbf{y}_2^T(t)]^T = [\mathbf{y}_{1,1}^T(t)\ldots\mathbf{y}_{1,J}^T(t)\mathbf{y}_{2,1}^T(t)\ldots\mathbf{y}_{2,J}^T(t)]^T$. Towards obtaining a first-order recursion for $\mathbf{y}(t)$, we require for all $j \in \mathcal{J}$:

(a1) Sensor observations obey $x_j(t) = \mathbf{h}_j^T(t)\mathbf{s}_0(t-1) + \epsilon_j(t)$, where the zero-mean white noise $\{\epsilon_j(t)\}$ has variance $\sigma_{\epsilon_j}^2$.

To concisely capture the effects of both observation and communication noise on the estimation errors across the WSN, define the $Jp \times 1$ noise vectors $\boldsymbol{\epsilon}(t) := 2\mu[\mathbf{h}_1^T(t)\epsilon_1(t)\ldots\mathbf{h}_J^T(t)\epsilon_J(t)]^T$ and $\bar{\boldsymbol{\eta}}(t) := [\bar{\boldsymbol{\eta}}_1^T(t)\ldots\bar{\boldsymbol{\eta}}_J^T(t)]^T$; where $\{\bar{\boldsymbol{\eta}}_j(t)\}_{j\in\mathcal{J}}$ are given by

$$\bar{\boldsymbol{\eta}}_j(t) := \sum_{b\in\mathcal{B}_j}\sum_{j'\in\mathcal{N}_b}\frac{\bar{\boldsymbol{\eta}}_b^{j'}(t)}{|\mathcal{N}_b|}. \quad (7)$$

and $\mathbf{R}_{\bar{\boldsymbol{\eta}}} := E[\bar{\boldsymbol{\eta}}(t)\bar{\boldsymbol{\eta}}^T(t)]$. Introduce the $p(\sum_{b\in\mathcal{B}}|\mathcal{N}_b|) \times 1$ vector

$$\boldsymbol{\eta}(t) := [\{(\boldsymbol{\eta}_{j'}^{b_1}(t))^T\}_{j'\in\mathcal{N}_{b_1}} \ldots \{(\boldsymbol{\eta}_{j'}^{b_{|\mathcal{B}|}}(t))^T\}_{j'\in\mathcal{N}_{b_{|\mathcal{B}|}}}]^T \quad (8)$$

comprising the receiver noise of the bridge sensors' transmissions to their neighbors, and define $\mathbf{R}_{\boldsymbol{\eta}} := E[\boldsymbol{\eta}(t)\boldsymbol{\eta}^T(t)]$. Finally, consider the $Jp \times Jp$ generalized "two-hop range" Laplacian matrix [2]

$$\mathbf{A} := \text{bdiag}(|\mathcal{B}_1|\mathbf{I}_p, \ldots, |\mathcal{B}_J|\mathbf{I}_p) - \sum_{b\in\mathcal{B}}\frac{(\mathbf{e}_b\otimes\mathbf{I}_p)(\mathbf{e}_b\otimes\mathbf{I}_p)^T}{|\mathcal{N}_b|} \quad (9)$$

where $\mathbf{e}_b$ represents the b-th column of the adjacency matrix $\mathbf{E}$. Based on these definitions, it is possible to state the main result of this section that will be instrumental in the subsequent analysis [7].

Lemma 1: *Under (a1) and for $t \geq 0$, the global state $\mathbf{y}(t)$ evolves according to*

$$\mathbf{y}(t+1) = bdiag(\mathbf{I}_{Jp}, c\mathbf{A})\mathbf{z}(t+1) + \begin{bmatrix} 2\mu c\mathbf{I}_{Jp} \\ -c\mathbf{I}_{Jp}\end{bmatrix}\bar{\boldsymbol{\eta}}(t) + \begin{bmatrix} 2\mu c\mathbf{P}_\alpha \\ -c\mathbf{P}_\alpha\end{bmatrix}\boldsymbol{\eta}(t) \quad (10)$$

where the inner state $\mathbf{z}(t) := [\mathbf{z}_1^T(t)\ \mathbf{z}_2^T(t)]^T$ *is arbitrarily initialized and updated according to*

$$\begin{aligned}\mathbf{z}(t+1) = \boldsymbol{\Phi}(t+1,\mu)\mathbf{z}(t) + \mathbf{R}_h^{\alpha}(t+1)\bar{\boldsymbol{\eta}}(t-1) \\ + \mathbf{R}_h^{\beta}(t+1)\boldsymbol{\eta}(t-1) + [\boldsymbol{\epsilon}^T(t+1) \quad \mathbf{0}^T]^T \\ - [\mathbf{1}_J^T \otimes (\mathbf{s}_0(t+1) - \mathbf{s}_0(t))^T \quad \mathbf{0}^T]^T, \end{aligned} \tag{11}$$

and $\boldsymbol{\Phi}(t,\mu)$ *consists of the blocks* $[\boldsymbol{\Phi}(t,\mu)]_{11} = \mathbf{I}_{Jp} - 2\mu(\mathbf{R}_h(t) + c\mathbf{A})$, $[\boldsymbol{\Phi}(t,\mu)]_{12} = -\mu c\mathbf{A}$ *and* $[\boldsymbol{\Phi}(t,\mu)]_{21} = [\boldsymbol{\Phi}(t,\mu)]_{22} = \mathbf{A}\mathbf{A}^{\dagger}$, *with* $\mathbf{R}_h(t) := bdiag\left(\mathbf{h}_1(t)\mathbf{h}_1^T(t), \ldots, \mathbf{h}_J(t)\mathbf{h}_J^T(t)\right)$. *The matrices* $\mathbf{R}_h^{\alpha}(t)$ *and* $\mathbf{R}_h^{\beta}(t)$ *are defined as*

$$\begin{aligned}\mathbf{R}_h^{\alpha}(t) &:= [\mu c\mathbf{I}_{Jp} - 4\mu^2 c(\mathbf{R}_h(t) + c\mathbf{A})^T, 2\mu c\mathbf{I}_{Jp}]^T \\ \mathbf{R}_h^{\beta}(t) &:= [\mu c(3\mathbf{P}_{\alpha} - 2\mathbf{P}_{\beta})^T - 4\mu^2 c((\mathbf{R}_h(t) + c\mathbf{A})\mathbf{P}_{\alpha})^T, \\ &\qquad 2\mu c\mathbf{P}_{\alpha}^T + 2\mathbf{C}_R^T]^T \end{aligned} \tag{12}$$

with $\mathbf{C}_R$ *chosen such that* $\mathbf{A}\mathbf{C}_R = \mathbf{P}_{\alpha} - \mathbf{P}_{\beta}$ *and the structure of the time-invariant matrices* $\mathbf{P}_{\alpha}$ *and* $\mathbf{P}_{\beta}$ *is given in Appendix A.*

3.2. Performance Metrics

When it comes to performance evaluation of adaptive filters, it is customary to consider the so called mean-square error (MSE), excess mean-square error (EMSE) and mean-square deviation (MSD) as figures of merit [3], [4]. In the present setup for distributed adaptive estimation, it is pertinent to address both global (network) and local (per-sensor) performance [1]. The aforementioned local quantities can be defined for all $j \in \mathcal{J}$ as

$$\begin{aligned}\text{MSE}_j(t) &:= E[e_j(t)^2], \\ \text{EMSE}_j(t) &:= E[(\mathbf{h}_j^T(t)\mathbf{y}_{1,j}(t-1))^2], \\ \text{MSD}_j(t) &:= E[\|\mathbf{y}_{1,j}(t)\|^2],\end{aligned}$$

whereas the global counterparts are defined as the respective averages across sensors, e.g., $\text{MSE}(t) := J^{-1}\sum_{j=1}^{J} E[e_j(t)^2]$ and so on. Next, note that in virtue of (a1) it is possible to write $e_j(t) = -\mathbf{h}_j^T(t)\mathbf{y}_{1,j}(t-1) + \epsilon_j(t)$ and also assume $\forall\, j \in \mathcal{J}$ that:
(a2) $\{\mathbf{h}_j(t)\}$ is white with covariance matrix $\mathbf{R}_{h_j} \succ \mathbf{0}$; and
(a3) $\{\mathbf{h}_j(t)\}, \{\epsilon_j(t)\}, \{\boldsymbol{\eta}_j^b(t)\}$ and $\{\bar{\boldsymbol{\eta}}_b^j(t)\}$ are independent.
Assumptions (a1)-(a3) comprise the widely adopted *independence setting*, which is instrumental in rendering the subsequent performance analysis tractable. Although (a2) can be grossly violated in, e.g., FIR filtering of signals (regressors) with a shift structure as in Section 2.1, for small step-sizes the upshot of the analysis does extend to correlated data (see also Remark 2).

Because $\mathbf{y}_{1,j}(t-1)$ is independent of the zero-mean $\{\mathbf{h}_j(t), \epsilon_j(t)\}$, from (a1)-(a3) ones finds that $\text{MSE}_j(t) = \text{EMSE}_j(t) + \sigma_{\epsilon_j}^2$; hence, it suffices to focus on the evaluation of $\text{EMSE}_j(t)$. Introducing the j-th local error covariance matrix $\mathbf{R}_{y_{1,j}}(t) := E[\mathbf{y}_{1,j}(t)\mathbf{y}_{1,j}^T(t)]$, then $\text{MSD}_j(t) = \text{tr}(\mathbf{R}_{y_{1,j}}(t))$ and under (a1)-(a3) a simple manipulation yields $\text{EMSE}_j(t) = \text{tr}(\mathbf{R}_{h_j}\mathbf{R}_{y_{1,j}}(t-1))$. Next, define $\mathbf{R}_y(t) := E[\mathbf{y}(t)\mathbf{y}(t)^T]$ and note that the global error covariance matrix corresponds to its $Jp \times Jp$ upper left submatrix $[\mathbf{R}_y(t)]_{11}$. Further, its j-th $p \times p$ diagonal submatrix ($j = 1, \ldots, J$) denoted by $[\mathbf{R}_y(t)]_{11,j}$ is $\mathbf{R}_{y_{1,j}}(t)$. It follows that with $\mathbf{R}_h := E[\mathbf{R}_h(t)] = \text{bdiag}(\mathbf{R}_{h_1}, \ldots, \mathbf{R}_{h_J})$, the global performance metrics are given by $\text{MSD}(t) = J^{-1}\text{tr}([\mathbf{R}_y(t)]_{11})$ and $\text{EMSE}(t) = J^{-1}\text{tr}(\mathbf{R}_h[\mathbf{R}_y(t-1)]_{11})$, which motivates deriving a closed-form expression for $\mathbf{R}_y(t)$.

4. STATIONARY CASE

Consider first a stationary setup by assuming that:
(a4) The true parameter is time-invariant, i.e., $\mathbf{s}_0(t) = \mathbf{s}_0$.
Under (a4) the last term in (11), due to parameter velocity, vanishes.

4.1. Mean Stability

From Lemma 1 it is straightforward to establish that D-LMS' local estimates are asymptotically unbiased, implying that consensus in the mean-sense is achieved on $\mathbf{s}_0$ [7].

Proposition 1: *Under (a1)-(a4), the D-LMS achieves consensus in the mean-sense, i.e.,* $\lim_{t\to\infty} E[\mathbf{y}_{1,j}(t)] = \mathbf{0}\ \forall\, j \in \mathcal{J}$ *provided the step-size is chosen such that* $\mu \in (0, \mu_u)$ *with*

$$\mu_u := 2\min(\lambda_{\max}^{-1}(\mathbf{R}_h + c\mathbf{A}), \lambda_{\max}^{-1}(2\mathbf{R}_h + (3c/2)\mathbf{A})). \tag{13}$$

4.2. MSE Stability and Performance Evaluation

Turning to MSE stability and performance analysis, observe from (10) that $\mathbf{y}_1(t+1) = \mathbf{z}_1(t+1) + 2\mu c[\bar{\boldsymbol{\eta}}(t) + \mathbf{P}_{\alpha}\boldsymbol{\eta}(t)]$. Under (a2)-(a3) $\mathbf{z}_1(t+1)$ is independent of the zero-mean $\{\bar{\boldsymbol{\eta}}(t), \boldsymbol{\eta}(t)\}$, so

$$[\mathbf{R}_y(t)]_{11} = [\mathbf{R}_z(t)]_{11} + 4(\mu c)^2[\mathbf{R}_{\bar{\boldsymbol{\eta}}} + \mathbf{P}_{\alpha}\mathbf{R}_{\boldsymbol{\eta}}\mathbf{P}_{\alpha}^T], \tag{14}$$

which readily prompts us to obtain $\mathbf{R}_z(t) = E[\mathbf{z}(t)\mathbf{z}^T(t)]$. Toward this end, observe that for all $j \in \mathcal{J}$ there exist unitary matrices $\mathbf{U}_j$ that we arrange in $\mathbf{U} = \text{bdiag}(\mathbf{U}_1, \ldots, \mathbf{U}_J)$ such that $\mathbf{U}_j\mathbf{R}_{h_j}\mathbf{U}_j^T = \boldsymbol{\Lambda}_j = \text{diag}(\lambda_1^j, \ldots, \lambda_p^j)$ and also $\mathbf{U}\mathbf{R}_h\mathbf{U}^T = \text{bdiag}(\boldsymbol{\Lambda}_1, \ldots, \boldsymbol{\Lambda}_J)$. For our subsequent arguments it will prove useful to introduce the (invertible) change of variables $\tilde{\mathbf{z}}(t) := \check{\mathbf{U}}\mathbf{z}(t)$ with $\check{\mathbf{U}} := \text{bdiag}(\mathbf{U}, \mathbf{U})$. Next, specialize (a2) by assuming:
(a5) $\{\mathbf{h}_j(t)\}$ is white Gaussian with covariance matrix $\mathbf{R}_{h_j} \succ \mathbf{0}$.
The Gaussianity assumption is instrumental in obtaining closed-form expressions for the regressors' fourth-order moments, which arise in the evaluation of $\mathbf{R}_{\tilde{z}}(t+1)$ [7].

Proposition 2: *Under (a3)-(a5) and for* $t \geq 0$, *the covariance matrix of* $\tilde{\mathbf{z}}(t)$ *obeys the first-order matrix recursion*

$$\begin{aligned}\mathbf{R}_{\tilde{z}}(t+1) = \mathcal{M}(\mathbf{R}_{\tilde{z}}(t), \tilde{\boldsymbol{\Phi}}(\mu), [\mathbf{R}_{\tilde{z}}(t)]_{11}) + \mathcal{M}(\mathbf{R}_{\bar{\boldsymbol{\eta}}}, \tilde{\mathbf{R}}_h^{\alpha}, 4\mu^2\mathbf{R}_{\bar{\boldsymbol{\eta}}}) \\ + \mathcal{M}(\mathbf{R}_{\boldsymbol{\eta}}, \tilde{\mathbf{R}}_h^{\beta}, 4\mu^2\mathbf{P}_{\alpha}\mathbf{R}_{\boldsymbol{\eta}}\mathbf{P}_{\alpha}^T) \\ + 4\mu^2 bdiag(\sigma_{\epsilon_1}^2\boldsymbol{\Lambda}_1, \ldots, \sigma_{\epsilon_J}^2\boldsymbol{\Lambda}_J, \mathbf{0})\end{aligned} \tag{15}$$

with $\tilde{\boldsymbol{\Phi}}(\mu) := \check{\mathbf{U}}\boldsymbol{\Phi}(\mu)\check{\mathbf{U}}^T$, $\tilde{\mathbf{R}}_h^i := \check{\mathbf{U}}E[\mathbf{R}_h^i(t)]$ *for* $i = \alpha, \beta$, *and*

$$\begin{aligned}\mathcal{M}(\mathbf{R}, \mathbf{S}, \mathbf{T}) := \mathbf{S}\mathbf{R}\mathbf{S}^T + 4\mu^2[bdiag((\mathbf{I}_J \otimes \mathbf{1}_{p\times p}) \circ \boldsymbol{\Lambda}\mathbf{T}\boldsymbol{\Lambda}, \mathbf{0}) \\ + bdiag(tr(\boldsymbol{\Lambda}_1[\mathbf{T}]_1)\boldsymbol{\Lambda}_1, \ldots, tr(\boldsymbol{\Lambda}_J[\mathbf{T}]_J)\boldsymbol{\Lambda}_J, \mathbf{0})]\end{aligned} \tag{16}$$

where $[\mathbf{T}]_j$ *stands for the* j*-th* $p \times p$ *diagonal sub-matrix of* $\mathbf{T}$.

Using (15), the transformation $\mathbf{R}_z(t) = \check{\mathbf{U}}^T\mathbf{R}_{\tilde{z}}(t)\check{\mathbf{U}}$ and (14) enables the closed-form evaluation of the $\text{MSE}(t)$, $\text{EMSE}(t)$ and $\text{MSD}(t)$ for all $t \geq 0$ by using the formulas in Section 3.2.

Remark 2 Stochastic averaging techniques [4, pg. 231] were applied in [2] to obtain an approximate global error covariance matrix $[\mathbf{R}_{\bar{y}}(t)]_{11}$ in the presence of temporally-correlated (non-) Gaussian data. A *trajectory locking* result stated therein further establishes that the approximation error vanishes as $\mu \to 0$. Interestingly, from the closed-form expression derived for $[\mathbf{R}_y(t)]_{11}$ under (a3)-(a5) one finds $\|[\mathbf{R}_y(t)]_{11} - [\mathbf{R}_{\bar{y}}(t)]_{11}\| = \mathcal{O}(\mu^2)$. By transitivity, the performance results of this paper are expected to accurately extend to general data settings, provided μ is sufficiently small.

The next step is to reformulate the matrix recursion (15) into a first-order vector recursion which is better suited for stability analysis. Specifically, we can vectorize (15) to obtain $\text{vec}[\mathbf{R}_{\tilde{z}}(t+1)] = \text{vec}[\mathcal{M}(\mathbf{R}_{\tilde{z}}(t), \tilde{\mathbf{\Phi}}(\mu), [\mathbf{R}_{\tilde{z}}(t)]_{11})] + \text{vec}[\mathbf{R}_{\tilde{\boldsymbol{\nu}}_S}]$, where $\mathbf{R}_{\tilde{\boldsymbol{\nu}}_S}$ stands for the (noise-induced) last three terms in the right-hand side (r.h.s.) of (15). Further simplification is possible by relying on properties of the vec[.] operator, as asserted by the following lemma established in [7].

Lemma 2: *Under (a3)-(a5) and for $t \geq 0$, the vectorized covariance matrix of $\tilde{\mathbf{z}}(t)$ obeys the first-order vector recursion*

$$\textit{vec}[\mathbf{R}_{\tilde{z}}(t+1)] = \tilde{\mathbf{\Psi}}(\mu)\textit{vec}[\mathbf{R}_{\tilde{z}}(t)] + \textit{vec}[\mathbf{R}_{\tilde{\boldsymbol{\nu}}_S}] \qquad (17)$$

where the $(2Jp)^2 \times (2Jp)^2$ transition matrix $\tilde{\mathbf{\Psi}}(\mu)$ is

$$\tilde{\mathbf{\Psi}}(\mu) := \tilde{\mathbf{\Phi}}(\mu) \otimes \tilde{\mathbf{\Phi}}(\mu) + 4\mu^2[(\textit{bdiag}(\mathbf{\Lambda}, \mathbf{0}) \otimes \textit{bdiag}(\mathbf{\Lambda}, \mathbf{0}))\times \textit{diag}(\textit{vec}[\mathbf{I}_{2Jp} \otimes \mathbf{1}_{p\times p}]) + \sum_{j=1}^{J} \mathbf{v}_j \mathbf{v}_j^T] \qquad (18)$$

and $\mathbf{v}_j := \textit{vec}[\textit{bdiag}(\mathbf{\Lambda} \circ [\textit{diag}(\mathbf{b}_{J,j}) \otimes \mathbf{1}_{p\times p}], \mathbf{0})] \ \forall\, j \in \mathcal{J}$.
An immediate consequence of Lemma 2 is that the D-LMS is MSE stable if and only if $|\lambda_{\max}(\tilde{\mathbf{\Psi}}(\mu))| < 1$. Although deriving explicit bounds on μ for stability appears intractable, the following proposition provides an important existence result [7].

Proposition 3: *Under (a1), (a3)-(a5) D-LMS algorithm is MSE stable provided that $\mu > 0$ is chosen sufficiently small.*

4.3. MSE Performance in Steady-State

Under the stability conditions on Proposition 2, the steady-state covariance matrix $\mathbf{R}_{\tilde{z}}(\infty) := \lim_{t\to\infty} \mathbf{R}_{\tilde{z}}(t)$ with bounded entries is guaranteed to exist. Indeed, $\text{vec}[\mathbf{R}_{\tilde{z}}(\infty)]$ can be obtained as fixed point of (17) so that solving yields

$$\text{vec}[\mathbf{R}_{\tilde{z}}(\infty)] = (\mathbf{I}_{(2Jp)^2} - \tilde{\mathbf{\Psi}}(\mu))^{-1}\text{vec}[\mathbf{R}_{\tilde{\boldsymbol{\nu}}_S}]. \qquad (19)$$

Note that if D-LMS is MSE stable, i.e., $\tilde{\mathbf{\Psi}}(\mu)$ is a stable matrix, from Gershgorin's circle Theorem $(\mathbf{I}_{(2Jp)^2} - \tilde{\mathbf{\Psi}}(\mu))^{-1}$ is guaranteed to exist. Exactly as before, all relevant local and global figures of merit in steady-state can be evaluated provided $[\mathbf{R}_y(\infty)]_{11}$ is available (c.f. Section 3.2). For that purpose, reshape (19) to obtain $\mathbf{R}_{\tilde{z}}(\infty)$, undo the change of variables to get $\mathbf{R}_z(\infty)$ and finally use (14).

5. TRACKING PERFORMANCE

Turning our attention to the nonstationary case and in conjunction with (a1) and (a2) [or (a5)], consider now $\forall\, j \in \mathcal{J}$:
(a6) Random-walk model, i.e., $\mathbf{s}_0(t) = \mathbf{s}_0(t-1) + \boldsymbol{\zeta}(t)$, where $\{\boldsymbol{\zeta}(t)\}$ is zero-mean white with covariance matrix $\mathbf{R}_\zeta \succ \mathbf{0}$; and
(a7) $\{\mathbf{h}_j(t)\}$, $\{\epsilon_j(t)\}$, $\{\boldsymbol{\zeta}(t)\}$, $\{\boldsymbol{\eta}_j^b(t)\}$, $\{\bar{\boldsymbol{\eta}}_b^j(t)\}$ are independent.
A random-walk is the simplest stochastic model to describe the variations of $\mathbf{s}_0(t)$. It could be arguably thought as not meaningful due to its increasing variance, thus violating the physical constraint $E[x_j^2(t)] < \infty$ for infinite time horizons. However, it is well justified as the resulting analysis sheds sufficient light on the key aspects of D-LMS when it comes to tracking.

Under (a6) and differently from the stationary case, the perturbation due to parameter velocity in (11) is not null but instead becomes $-[\mathbf{1}_J^T \otimes \boldsymbol{\zeta}^T(t+1)\ \mathbf{0}]^T$. However, from (a6) it does vanish under expectation so that as in classical LMS [4], the mean-stability analysis carries over, as asserted next.

Corollary 1: *Under (a1)-(a2), (a6)-(a7) D-LMS achieves consensus in the mean-sense, provided $\mu \in (0, \mu_u)$ with μ_u as in (13).*

5.1. MSE Stability and Performance Evaluation

The perturbation $-[\mathbf{1}_J^T \otimes \boldsymbol{\zeta}^T(t+1)\ \mathbf{0}]^T$ arising in the nonstationary setting is independent of all other noise quantities in virtue of (a7) and also of $\mathbf{z}(t)$ in (11) due to (a6)-(a7). Therefore, the tracking performance evaluation of D-LMS under the random-walk model can be carried out with minor efforts by extending the results in Section 4. Introducing the same change of variables $\tilde{\mathbf{z}}(t) := \check{\mathbf{U}}\mathbf{z}(t)$, it is possible to derive an exact closed-form recursion for $\mathbf{R}_{\tilde{z}}(t)$ [7].

Proposition 4: *Under (a5)-(a7) and for $t \geq 0$, the covariance matrix of $\tilde{\mathbf{z}}(t)$ obeys the first-order matrix recursion*

$$\begin{aligned}\mathbf{R}_{\tilde{z}}(t+1) &= \mathcal{M}(\mathbf{R}_{\tilde{z}}(t), \tilde{\mathbf{\Phi}}(\mu), [\mathbf{R}_{\tilde{z}}(t)]_{11}) \\ &+ \mathcal{M}(\mathbf{R}_{\bar{\boldsymbol{\eta}}}, \tilde{\mathbf{R}}_h^\alpha, 4\mu^2\mathbf{R}_{\bar{\boldsymbol{\eta}}}) + \mathcal{M}(\mathbf{R}_{\boldsymbol{\eta}}, \tilde{\mathbf{R}}_h^\beta, 4\mu^2\mathbf{P}_\alpha\mathbf{R}_{\boldsymbol{\eta}}\mathbf{P}_\alpha^T) \\ &+ 4\mu^2\textit{bdiag}(\sigma_{\epsilon_1}^2\mathbf{\Lambda}_1, \ldots, \sigma_{\epsilon_J}^2\mathbf{\Lambda}_J, \mathbf{0}) + \check{\mathbf{U}}\textit{bdiag}(\mathbf{1}_{J\times J} \otimes \mathbf{R}_\zeta, \mathbf{0})\check{\mathbf{U}}^T\end{aligned} \qquad (20)$$

and the equivalent first-order vector recursion after vectorization

$$\textit{vec}[\mathbf{R}_{\tilde{z}}(t+1)] = \tilde{\mathbf{\Psi}}(\mu)\textit{vec}[\mathbf{R}_{\tilde{z}}(t)] + \textit{vec}[\mathbf{R}_{\tilde{\boldsymbol{\nu}}_{NS}}] \qquad (21)$$

where $\mathbf{R}_{\tilde{\boldsymbol{\nu}}_{NS}}$ stands for the last four terms in the r.h.s. of (20) and $\tilde{\mathbf{\Psi}}(\mu)$ is defined in (18).
By inspecting (21) one realizes that the only difference with (17) stems from the constant forcing vector. Consequently, MSE stability conditions remain the same as in the stationary case.

Corollary 2: *Under (a1), and (a5)-(a7) the D-LMS algorithm is MSE stable provided that $\mu > 0$ is chosen sufficiently small.*

The steady-state tracking performance of D-LMS can be evaluated by repeating the steps described in Section 4.3, after replacing $\mathbf{R}_{\tilde{\boldsymbol{\nu}}_S}$ with $\mathbf{R}_{\tilde{\boldsymbol{\nu}}_{NS}}$ on the r.h.s. of (19).

5.2. Step-size Optimization

While MSE stability ensures, e.g., a bounded EMSE(∞); satisfactory tracking of $\mathbf{s}_0(t)$ ultimately requires the error to be small. This will depend on μ and the speed of parameter variation roughly dictated by $\text{tr}(\mathbf{1}_{J\times J} \otimes \mathbf{R}_\zeta)$. Interestingly, whenever $\text{tr}(\mathbf{1}_{J\times J} \otimes \mathbf{R}_\zeta)$ is comparable to $\text{tr}(\mathbf{R}_{\boldsymbol{\nu}_S})$ there exists an optimal $\mu^\star$ minimizing EMSE(∞). This should not be surprising since excessive adaptation leads to the same MSE inflation as in the absence of parameter variation, while if μ is too small the tracking ability may be lost and once again an MSE penalty is expected. Recall that

$$\begin{aligned}\text{EMSE}(\infty) &= \frac{1}{J}\sum_{j=1}^{J}\text{tr}(\mathbf{\Lambda}_j\mathbf{R}_{\tilde{z}_{1,j}}(\infty)) = \frac{1}{J}\sum_{j=1}^{J}\mathbf{v}_j^T\text{vec}[\mathbf{R}_{\tilde{z}}(\infty))] \\ &= \frac{1}{J}\sum_{j=1}^{J}\mathbf{v}_j^T(\mathbf{I}_{(2Jp)^2} - \tilde{\mathbf{\Psi}}(\mu))^{-1}\text{vec}[\mathbf{R}_{\tilde{\boldsymbol{\nu}}_{NS}}].\end{aligned} \qquad (22)$$

where in obtaining the second equality we used $\text{tr}(\mathbf{R}^T\mathbf{S}) = \text{vec}[\mathbf{R}]^T\text{vec}[\mathbf{S}]$ and the $\{\mathbf{v}_j\}$ were defined in Lemma 2. Now, $\mathbf{R}_{\tilde{\boldsymbol{\nu}}_{NS}} = \mathbf{R}_{\tilde{\boldsymbol{\nu}}_S} + \check{\mathbf{U}}\text{bdiag}(\mathbf{1}_{J\times J} \otimes \mathbf{R}_\zeta, \mathbf{0})\check{\mathbf{U}}^T$, where the aggregate noise covariance is $\mathbf{R}_{\tilde{\boldsymbol{\nu}}_S} = \mathcal{O}(\mu^2)$ [cf. (20),(12)] and the second summand is $\mathcal{O}(1)$. Roughly, $(\mathbf{I}_{(2Jp)^2} - \tilde{\mathbf{\Psi}}(\mu))^{-1} = \mathcal{O}(\mu^{-1})$ so that

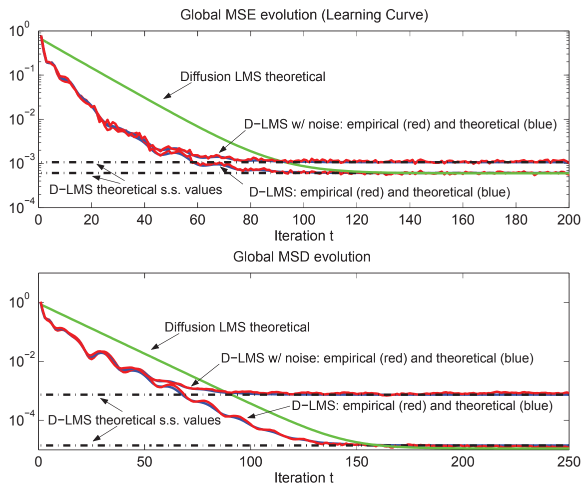

Fig. 1. Global performance evaluation for the stationary case.

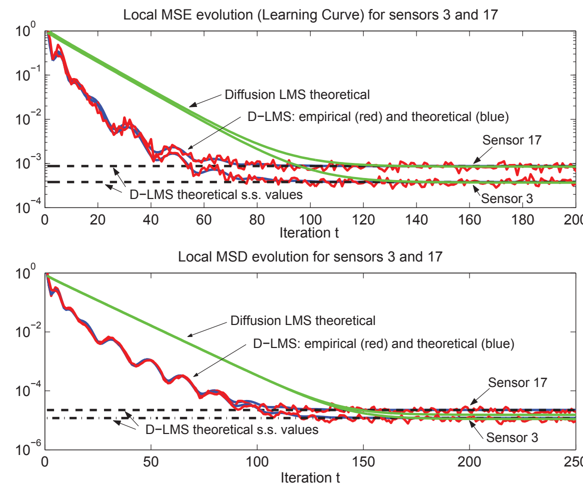

Fig. 2. Local performance evaluation for the stationary case.

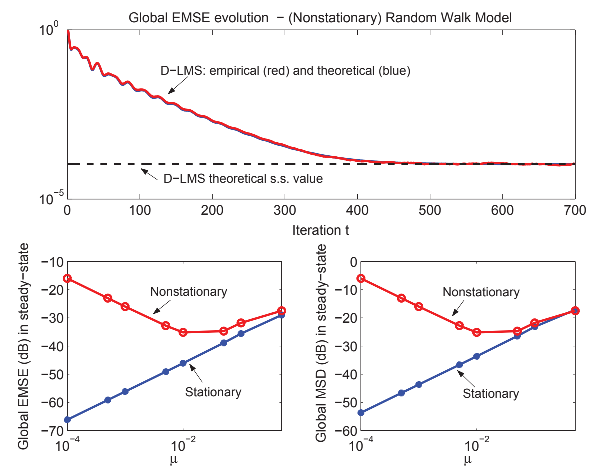

Fig. 3. (top) Global performance evaluation for the nonstationary case; (bottom) Steady-state EMSE and MSD values versus μ.

one finds from (22) that EMSE$(\infty) = \mathcal{O}(\mu^{-1})$ for small μ, whereas EMSE$(\infty) = \mathcal{O}(\mu)$ for moderate- to large values of the step-size approaching the stability bound. If e.g., tr$(\mathbf{1}_{J\times J}\otimes\mathbf{R}_\zeta) \gg$ tr$(\mathbf{R}_{\boldsymbol{\nu}_S})$ then $\mathbf{R}_{\tilde{\boldsymbol{\nu}}_{NS}} \approx \breve{\mathbf{U}}\text{bdiag}(\mathbf{1}_{J\times J}\otimes\mathbf{R}_\zeta, \mathbf{0})\breve{\mathbf{U}}^T$ in the whole range of stable step-sizes so that EMSE(∞) will not attain a minimum.

Unfortunately, deriving an explicit formula for $\mu^\star$ is a formidable task. If needed however, $1-$D minimization can be carried out numerically using, e.g., Newton's method, as the derivatives of the EMSE(∞) cost are readily computable in closed form.

6. NUMERICAL TESTS

With $J = 20$ sensors, a WSN is generated as a planar random geometric graph with communication range $r = 0.6$. For the examples with noisy links receiver AWGN with variance $\sigma_\eta^2 = 10^{-3}$ is added. A linear model [cf. (a1)] is adopted for the sensor data, with observation WGN of spatial variance profile $\sigma_{\epsilon_j}^2 = 10^{-3}\alpha_j$, with i.i.d. $\alpha_j \sim \mathcal{U}[0,1]$ (uniform distribution). For the different test cases and $p = 4$, the signal vector will either be $\mathbf{s}_0 = \mathbf{1}_p$ as in (a4), or adhere to the random-walk model in (a6) with $\mathbf{R}_\zeta = 10^{-6}\mathbf{I}_p$. The regressors $\mathbf{h}_j(t) = [h_j(t)\ldots h_j(t-p+1)]^T$ have entries which evolve according to $h_j(t) = (1-\rho)\beta_j h_j(t-1) + \sqrt{\rho}\nu_j(t)$ for all $j \in \mathcal{J}$. We choose $\rho = 7\times 10^{-1}$, the $\beta_j \sim \mathcal{U}[0,1]$ are i.i.d. in space, and the driving white noise $\nu_j(t) \sim \mathcal{U}[-\sqrt{3}\sigma_{\nu_j}, \sqrt{3}\sigma_{\nu_j}]$ has a spatial variance profile given by $\sigma_{\nu_j}^2 = 10^{-1}\gamma_j$ with $\gamma_j \sim \mathcal{U}[0,1]$ and i.i.d.. Thus, data is temporally-correlated and non-Gaussian.

With $\mu = 5\times 10^{-2}$ and $c = 1$, Fig. 1 depicts the global D-LMS performance through the evolution of the MSE(t) (learning curve) and MSD(t) figures of merit. Even though the data does not adhere to (a3)-(a5), the empirical curves (obtained via sample averaging 100 runs of D-LMS, i.e., recursions (4)-(6)) closely follow the theoretical trajectories predicted by Proposition 2. The steady-state limiting values in Section 4.3 are also extremely accurate. As suggested by intuition and corroborated by (14)-(15), the performance penalty due to non-ideal links is also apparent. Theoretical curves for the diffusion LMS [1, eqs. (73)-(74)] with Metropolis weights are also included. While in this case diffusion LMS has a slight edge on steady-state performance, note that it comes at the price of a much slower convergence rate. Similar overall conclusions can be drawn from Fig. 2, which shows the local performance of two randomly selected sensors.

With regards to D-LMS tracking performance, Fig. 3 (top) illustrates the global EMSE(t) evolution when $\mu = 5\times 10^{-2}$ and $c = 1$. Once more, it is remarkable how well the theoretical findings in Section 5.1 agree with the true behavior for all $t \geq 0$. The bottom plots in Fig. 3 corroborate the conclusions in Section 5.2, by showing the theoretically predicted dependence on μ of the steady-state global quantities EMSE(∞) and MSD(∞), for both the stationary and nonstationary setups described earlier. While the trend is similar for moderate- to large step-sizes, for small μ the MSE penalty in the tracking setup due to lack of adaptation becomes dominant, and is increasingly severe as $\mu \to 0$. The existence of $\mu^\star$ is also highlighted by Fig. 3.

7. CONCLUDING REMARKS

A detailed MSE performance analysis was conducted for D-LMS, both in the absence of parameter variation and when the parameter fluctuations adhere to a random-walk model. By deriving an exact matrix recursion for the global error covariance under the white Gaussian setting, the network and per-sensor performance figures of merit became available for any t, and in particular as $t \to \infty$. The tracking analysis led to the conclusion that – differently from the time-invariant case whereby one should decrease μ to reduce the steady-state error – for a time-varying parameter there exists an optimal $\mu^\star$, since a vanishing step-size renders D-LMS incapable of adapting to the variations. Numerical simulations corroborated the theoretical findings of this paper carry over to more pragmatic setups, including temporally-correlated (non-) Gaussian sensor data[1].

A. STRUCTURE OF MATRICES $\mathbf{P}_\alpha$ AND $\mathbf{P}_\beta$

From [2, Appendix D] $\mathbf{P}_\alpha := [\mathbf{p}_1 \dots \mathbf{p}_J]^T$ and $\mathbf{P}_\beta := [\mathbf{p}'_1 \dots \mathbf{p}'_J]^T$, where the $p(\sum_{b\in\mathcal{B}} |\mathcal{N}_b|) \times p$ submatrices $\mathbf{p}_j$, $\mathbf{p}'_j$ are given by $\mathbf{p}_j := [(\mathbf{p}_{j,1})^T \dots (\mathbf{p}_{j,|\mathcal{B}|})^T]^T$ and $\mathbf{p}'_j := [(\mathbf{p}'_{j,1})^T \dots (\mathbf{p}'_{j,|\mathcal{B}|})^T]^T$, with $\mathbf{p}_{j,r}, \mathbf{p}_{j',r}$ defined for $r = 1, \dots, |\mathcal{B}|$ as

$$\mathbf{p}_{j,r}^T := \begin{cases} \mathbf{b}^T_{|\mathcal{N}_{b_r}|,r(j)} \otimes \mathbf{I}_p & \text{if } j \in \mathcal{N}_{b_r} \\ \mathbf{0}_{p\times|\mathcal{N}_{b_r}|p} & \text{if } j \notin \mathcal{N}_{b_r} \end{cases},$$

$$(\mathbf{p}'_{j,r})^T := \begin{cases} |\mathcal{N}_{b_r}|^{-1}\mathbf{1}_{1\times|\mathcal{N}_{b_r}|} \otimes \mathbf{I}_p & \text{if } j \in \mathcal{N}_{b_r} \\ \mathbf{0}_{p\times|\mathcal{N}_{b_r}|p} & \text{if } j \notin \mathcal{N}_{b_r} \end{cases}.$$

Note that $r(j) \in \{1, \dots, |\mathcal{N}_{b_r}|\}$ denotes the order in which $\boldsymbol{\eta}_j^{b_r}(t)$ appears in $\{\boldsymbol{\eta}_{j'}^{b_r}(t)\}_{j'\in\mathcal{N}_{b_r}}$ [cf. (8)].

B. REFERENCES

[1] C. G. Lopes and A. H. Sayed, "Diffusion least-mean squares over adaptive networks: Formulation and performance analysis," *IEEE Trans. on Signal Processing*, vol. 56, pp. 3122–3136, July 2008.

[2] I. D. Schizas, G. Mateos, and G. B. Giannakis, "Distributed LMS for consesus-based in-network adaptive processing," *IEEE Trans. on Signal Processing*, 2008 (revised).

[3] A. H. Sayed, *Adaptive Filtering*, John Wiley & Sons, 2008.

[4] V. Solo and X. Kong, *Adaptive Signal Processing Algorithms: Stability and Performance*, Prentice Hall, 1995.

[5] P. Stoica and R. Moses, *Spectral Analysis of Signals*, Prentice Hall, 2005.

[6] D. P. Bertsekas and J. N. Tsitsiklis, *Parallel and Distributed Computation: Numerical Methods*, Athena-Scientific, 1999.

[7] G. Mateos, I. D. Schizas, and G. B. Giannakis, "Performance analysis of the consensus-based LMS algorithm," *IEEE Trans. on Signal Processing*, 2008 (in preparation).

[1] The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies of the Army Research Laboratory or the U. S. Government.