Distributed LMS for Consensus-Based In-Network Adaptive Processing

Ioannis D. Schizas, *Student Member, IEEE*, Gonzalo Mateos, *Student Member, IEEE*, and Georgios B. Giannakis, *Fellow, IEEE*

***Abstract*—Adaptive algorithms based on in-network processing of distributed observations are well-motivated for online parameter estimation and tracking of (non)stationary signals using ad hoc wireless sensor networks (WSNs). To this end, a fully *distributed* least mean-square (D-LMS) algorithm is developed in this paper, offering simplicity and flexibility while solely requiring single-hop communications among sensors. The resultant estimator minimizes a pertinent squared-error cost by resorting to i) the alternating-direction method of multipliers so as to gain the desired degree of parallelization and ii) a stochastic approximation iteration to cope with the time-varying statistics of the process under consideration. Information is efficiently percolated across the WSN using a subset of "bridge" sensors, which further tradeoff communication cost for robustness to sensor failures. For a linear data model and under mild assumptions aligned with those considered in the centralized LMS, stability of the novel D-LMS algorithm is established to guarantee that local sensor estimation error norms remain bounded most of the time. Interestingly, this weak stochastic stability result extends to the pragmatic setup where intersensor communications are corrupted by additive noise. In the absence of observation and communication noise, consensus is achieved almost surely as local estimates are shown exponentially convergent to the parameter of interest with probability one. Mean-square error performance of D-LMS is also assessed. Numerical simulations: i) illustrate that D-LMS outperforms existing alternatives that rely either on information diffusion among neighboring sensors, or, local sensor filtering; ii) highlight its tracking capabilities; and iii) corroborate the stability and performance analysis results.**

***Index Terms*—Distributed estimation, LMS algorithm, wireless sensor networks (WSNs).**

I. Introduction

DRIVEN by a wide span of foreseen applications, decentralized estimation of signals based on observations acquired by spatially distributed sensors has attracted much attention recently. Deployment of ad hoc wireless sensor networks (WSNs) based on single-hop communications is envisioned to perform various adaptive signal processing tasks, including distributed noise cancellation, power spectrum estimation, localization, field monitoring, and target tracking. Different from WSN topologies that include a fusion center (FC), ad hoc ones have to rely on *in-network* processing. The absence of a central processing unit prompts local sensor estimates to eventually consent to a common global estimate while fully exploiting spatial correlations to maximize estimation performance.

Several noteworthy contributions have built up the field of consensus-based distributed estimation. Achieving consensus across agents was considered in vehicle coordination [5], as well as in distributed sample-averaging of sensor observations [12], [26]. A general distributed estimation framework was put forth in [17] and [18]. In the aforementioned schemes, sensors acquire data only once and then locally exchange messages to reach consensus. Extensions for distributed tracking of the sample-average of time-varying signals can be found in, e.g., [13] and [22]. Sequential in-time incorporation of sensor observations to enrich the estimation process was considered in [27], in the context of linear least-squares parameter estimation. The space–time diffusion algorithm of [27] requires knowledge of the data model and costly exchanges of matrices among neighbors, while the requirement for diminishing step-sizes renders it incapable of tracking time-varying signals. Distributed Kalman filtering approaches have been also reported [11], [17], but they are applicable when the state and observation models are known.

In many applications, however, sensors need to perform estimation in a constantly changing environment without having available a (statistical) model for the underlying processes of interest. This motivates the development of distributed adaptive estimation schemes. The first such approach introduced a sequential scheme whereby LMS-type adaptive filtering per sensor allows the network to account for time variations in the signal statistics [8]. For more general estimators, a similar stochastic incremental gradient descent algorithm was developed in [15], which subsumes [8] as a special case. The incremental LMS schemes in [8], [15] may outperform a centralized implementation of LMS in terms of convergence rate and steady-state error, while entailing a relatively low communication overhead. These features make them appealing, especially for small-size WSNs. However, such schemes inherently require a Hamiltonian cycle through which signal estimates are sequentially circulated from sensor to sensor. In the eventuality of a sensor failure, determination of a new cycle is an NP-hard problem [14], thus challenging the applicability of incremental schemes in medium- to large-size WSNs. Avoiding the need of such a cycle and increasing the degree of collaboration among

Manuscript received January 17, 2008; accepted January 06, 2009. First published February 24, 2009; current version published May 15, 2009. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Aleksandar Dogandzic. Work in this paper was supported by the USDoD ARO Grant 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. Part of the paper was presented at the Forty-Fifth Annual Allerton Conference on Communication, Control and Computing, Monticello, IL, September 26–28, 2007 and at the International Conference on Acoustics, Speech and Signal Processing, Las Vegas, NV, March 30–April 4, 2008.

The authors are with the Department of Electrical and Computer Engineering, University of Minnesota, Minneapolis, MN 55455 USA (e-mail: schiz001@umn.edu; mate0058@umn.edu; georgios@umn.edu).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TSP.2009.2016226

neighbors, the so-termed diffusion LMS [9] offers an improved alternative at the price of increasing communication cost.

The present paper develops a consensus-based distributed (D-)LMS algorithm for in-network adaptive processing using ad hoc WSNs with noisy links, which circumvents the requirement of a cycle. Its simplicity matches well the scarcity of communication and computation resources characterizing WSNs. In contrast with [9], the algorithm is derived from a well-posed estimation criterion optimized using the alternating-direction method of multipliers (AD-MoM) and stochastic approximation techniques. Different from [8], [9], and [15], the novel D-LMS scheme accounts for intersensor communication noise, in which setup local estimation errors are shown to be stochastically bounded. In the absence of noise, the local estimates obtained via D-LMS converge exponentially fast to the true parameters. These stability properties, also present in the classical LMS algorithm (see, e.g., [20]), are established without invoking the independence and Gaussianity conditions assumed in [8]–[10]. Stochastic averaging arguments are further utilized to approximate the MSE associated with D-LMS. Moreover, D-LMS is shown flexible to tradeoff communication cost for robustness to sensor failures by imposing consensus only within a subset of the available sensors. The optimization setup used here to derive the distributed adaptive algorithm resembles the one in [17] and [18]. However, as in [8], [9], and [15], D-LMS offers novel and attractive features, including i) online incorporation and processing of new data across sensors and ii) a distributed adaptive estimation scheme for applications, where a statistical model of variations is not available (this is needed in, e.g., [11] and [17]).

In Section II, we introduce the WSN model and the optimization problem defining the desired estimator. Building on [18], we recast the original formulation into an equivalent constrained optimization problem, whose solution becomes available in a distributed fashion using the AD-MoM and stochastic approximation iterations leading to the novel D-LMS (Section III-A). Next, we describe its operation and required communications, and further elaborate on the intuition and flexibility of the resulting algorithm (Section III-B), before demonstrating its merits via numerical simulations in Section III-C. Turning our attention to performance analysis, the challenging problems of stochastic stability and asymptotic MSE characterization are addressed in Sections IV-A and IV-B. Concluding remarks are given in Section V.

Notation: Bold uppercase letters will denote matrices with ij-th entry $[\cdot]_{ij}$, whereas bold lowercase letters will stand for column vectors (ith entry denoted by $[\cdot]_i$); operators $\otimes$, $(\cdot)^T$, $(\cdot)^\dagger$, $\lambda_{\max}(\cdot)$, $\mathrm{diag}(\cdot)$, $\mathrm{bdiag}(\cdot)$, and $E[\cdot]$ will denote Kronecker product, transposition, matrix pseudoinverse, spectral radius, diagonal matrix (arguments are scalar diagonal entries), block diagonal matrix (arguments are matrix diagonal entries), and expectation, respectively. For both vector and matrices, $\|\cdot\|$ will stand for the 2-norm and $|\cdot|$ for the cardinality of a set or the magnitude of a scalar. 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.

II. Preliminaries and Problem Statement

Consider an ad hoc WSN comprising J sensors where only single-hop communications are allowed, i.e., sensor j can only communicate with sensors in its neighborhood $\mathcal{N}_j \subseteq \{1,\ldots,J\} := \mathcal{J}$, with the convention $j \in \mathcal{N}_j$. Assuming that intersensor links are symmetric, the WSN is modeled as an undirected graph whose vertices are the sensors and its edges represent the available links. Global connectivity information is captured by the symmetric adjacency matrix $\mathbf{E} \in \mathbb{R}^{J\times J}$, where $[\mathbf{E}]_{ij} = 1$ if $i \in \mathcal{N}_j$ and $[\mathbf{E}]_{ij} = 0$ otherwise. This model includes the widely adopted planar random geometric graph $\mathcal{G}^2(J,r)$ [4], where J sensors are randomly placed over the unity square, while connectivity of two nodes is ensured so long as their Euclidean distance is less than a pre-specified communication range r. To ensure that the data from an arbitrary sensor can eventually percolate through the entire network, the following is assumed:

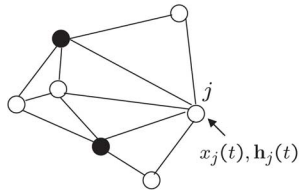

Fig. 1. Ad hoc wireless sensor network (WSN) with bridge sensors.

a1) *The WSN graph is connected; i.e., there exists a (possibly) multihop communication path connecting any two sensors.*

Different from [8], [9], and [15], the present network model accounts explicitly for nonideal sensor-to-sensor links, through a zero-mean additive noise vector $\boldsymbol{\eta}_j^i(t)$ with covariance matrix $\mathbf{R}_{\eta_{j,i}} := E\left[\boldsymbol{\eta}_j^i(t)\boldsymbol{\eta}_j^i(t)^T\right]$ corrupting signals received at sensor j from sensor i at discrete-time instant t. The noise vectors $\left\{\boldsymbol{\eta}_j^i(t)\right\}_{j\in\mathcal{J}}^{i\in\mathcal{N}_j}$ are assumed temporally and spatially uncorrelated. Because the results in this paper do not depend on the noise pdf, this model incorporates, but is not limited to receiver additive white Gaussian noise (AWGN). A sample ad hoc WSN is depicted in Fig. 1.

The WSN is deployed to estimate a signal vector $\mathbf{s}_0 \in \mathbb{R}^{p\times 1}$. 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 acquires a scalar observation $x_j(t)$, both assumed zero mean without loss of generality. A similar data setting was considered also in [8] and [9]. Introducing the global vector $\mathbf{x}(t) := [x_1(t)\ldots x_J(t)]^T \in \mathbb{R}^{J\times 1}$ and matrix $\mathbf{H}(t) := [\mathbf{h}_1(t)\ldots\mathbf{h}_J(t)]^T \in \mathbb{R}^{J\times p}$, the global LMS estimator of interest can be written as [8], [16, p. 49], [21, p. 14]

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

For jointly stationary $\{\mathbf{x}(t), \mathbf{H}(t)\}$, solving (1) leads to the well-known Wiener filter estimate $\hat{\mathbf{s}}_W = \mathbf{R}_H^{-1}\mathbf{r}_{Hx}$, where $\mathbf{R}_H := E[\mathbf{H}^T(t)\mathbf{H}(t)]$ and $\mathbf{r}_{Hx} := E[\mathbf{H}^T(t)\mathbf{x}(t)]$; see, e.g., [21, p. 15]. If $\mathbf{R}_H$ and $\mathbf{r}_{Hx}$ were known, then a steepest-descent iteration

$$\hat{\mathbf{s}}(t) = \hat{\mathbf{s}}(t-1) + \mu[\mathbf{r}_{Hx} - \mathbf{R}_H\hat{\mathbf{s}}(t-1)]$$

with sufficiently small step-size μ would converge to $\hat{\mathbf{s}}_W$ while avoiding the burden of inverting $\mathbf{R}_H$. In many linear regression applications involving online processing of data, this covariance

information may be either unavailable or time varying, and thus impossible to update continuously. Targeting low complexity implementations, one often resorts to the centralized (C-) LMS algorithm; see, e.g., [21, p. 77]

$$\hat{\mathbf{s}}(t)=\hat{\mathbf{s}}(t-1)+\mu\mathbf{H}^T(t)\left[\mathbf{x}(t)-\mathbf{H}(t)\hat{\mathbf{s}}(t-1)\right] \quad (2)$$

which relies on $\mathbf{R}_H \approx \mathbf{H}^T(t)\mathbf{H}(t)$ and $\mathbf{r}_{Hx} \approx \mathbf{H}^T(t)\mathbf{x}(t)$ to coarsely approximate the ensemble averages instantaneously. Considering a constant step-size μ, in order to allow for tracking of a possibly time-varying $\mathbf{s}_0(t)$, the C-LMS algorithm yields stochastic iterates $\hat{\mathbf{s}}(t)$ that do not converge to, but hover around, the desired signal of interest. Stability analysis of C-LMS has a long history. General results can be found in [7], [16], and [20] that also include surveys of prior art. One of the main results reported in [20] is that for observations adhering to a linear model, i.e., $\mathbf{x}(t)=\mathbf{H}(t)\mathbf{s}_0+\boldsymbol{\epsilon}(t)$ with $\mathbf{H}(t)$ assumed stationary ergodic with finite fourth-order moments and $\mathbf{R}_H > \mathbf{0}$, recursion (2) with sufficiently small step-size μ provably i) yields an estimation error whose norm remains most of the time within a finite interval, i.e., $\lim_{\delta\to\infty}\sup_{t\geq 0}\Pr\left[\|\hat{\mathbf{s}}(t)-\mathbf{s}_0\|\geq\delta\right]=0$ even in the presence of noise and ii) provides estimates that are almost surely (a.s.) convergent to the true parameter $\mathbf{s}_0$ at an exponential rate in the absence of observation noise. The stability notion described in i) is referred to as *weak stochastic stability*, and estimation errors are said to be weakly stochastic bounded (WSB) [20]. Note that in ii) a.s. convergence is with respect to the probability measure induced by the random regressors $\{\mathbf{h}_j(t)\}_{j=1}^J$.

Remark 1 (Application to Distributed Linear Regression): An interesting application where the need for linear regression arises is spectrum estimation. Specifically, suppose sensors observe a narrowband source to determine its spectral peaks, which can assist them disclose hidden periodicities due to a physical phenomenon controlled by, e.g., a natural heat source. The source of interest propagates through multipath channels and is contaminated with additive noise when sensed at the sensors. The unknown source-sensor channels may introduce deep fades at the frequency band occupied by the source. Thus, having each sensor operating on its own may lead to faulty assessments. The available spatial diversity to effect improved spectral estimates can only be achieved via sensor collaboration.

Let $\theta(t)$ denote the narrowband source of interest, which can be modeled as an autoregressive (AR) process [23, p. 106]

$$\theta(t)=-\sum_{\tau=1}^{p}\alpha_\tau\theta(t-\tau)+w(t) \quad (3)$$

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 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 $\{c_{j,l}\}$ and is contaminated with additive sensing noise $\bar{\epsilon}_j(t)$ to yield the observation

$$x_j(t)=\sum_{l=0}^{L_j-1}c_{j,l}\theta(t-l)+\bar{\epsilon}_j(t). \quad (4)$$

Since $x_j(t)$ is an autoregressive moving average (ARMA) process, it can be written as [23]

$$x_j(t)=-\sum_{\tau=1}^{p}\alpha_\tau x_j(t-\tau)+\sum_{\tau'=1}^{m}\beta_{\tau'}\tilde{\eta}_j(t-\tau'),\quad j\in\mathcal{J} \quad (5)$$

where the MA coefficients $\{\beta_{\tau'}\}$ and the variance of the white noise process $\tilde{\eta}_j(t)$ 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 (5) can be treated as observation noise, i.e., $\epsilon_j(t):=\sum_{\tau'=1}^{m}\beta_{\tau'}\tilde{\eta}_j(t-\tau')$. This is very 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 we let $\mathbf{s}_0:=[\alpha_1\ldots\alpha_p]^T$. From (5) 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. Distributed spectrum estimation has been considered also in [6] utilizing generalized projection schemes. Assumptions in [6] include ideal any-to-any communications and known source-sensor channels.

For different estimation/tracking applications suitable reformulation may be needed in order to acquire linear regressors based on the available information across sensors. For example, in target tracking applications where sensors rely on power or range measurements, the nonlinear data models must be linearized before obtaining regressors as a function of sensor observations; see, e.g., [1, p. 137]. Another possibility is to obtain the regression vectors from the physics of the problem, using standard kinematic models with constant velocity or acceleration that are well documented in the tracking literature; see, e.g., [1, Ch. 6].

Remark 2 (Motivation for In-Network Processing): Both C-LMS and incremental LMS [8] provide comparable performance benchmarks for distributed LMS-type adaptation rules, as every update encompasses all the information available in the network. Although both the observations $\mathbf{x}(t)$ and regressor rows in $\mathbf{H}(t)$ are actually disseminated across the WSN, in the broad context of sensor network processing one could envision an implementation of the C-LMS using an FC-based topology. This, however, comes at the price of isolating the network's point of failure and may challenge communications as the WSN scales over a larger geographic area, since far away sensors will require higher power to reach the FC, thus diminishing their battery lifetime.

In the context of Remarks 1 and 2, this paper aims to develop and analyze in terms of stability and performance, a fully *distributed* (D-) LMS algorithm for in-network adaptive processing using ad hoc WSNs. In a nutshell, the described setup naturally suggests three characteristics that the algorithm should exhibit: i) stability properties analogous to C-LMS, ii) processing at the sensor level should be kept as simple as possible; and iii) communications among sensors should be confined to single-hop exchanges.

III. The D-LMS Algorithm

In this section, we introduce the D-LMS algorithm, first going through the algorithm construction and salient features of its operation. The approach followed includes three main building blocks: i) recast (1) into an equivalent form amenable to distributed implementation, ii) split the optimization problem into smaller and simpler subtasks executed locally at each sensor, and iii) invoke a stochastic approximation iteration to obtain an adaptive LMS-type of algorithm that can both handle the unavailability/variation of statistical information, and also remain robust to signal variations. We further interpret the resulting D-LMS recursions to gain insights on how local and net-

work-wide information are combined in the learning process, and build intuition on the mechanisms employed to reach consensus among sensors on the adaptive estimate.

To distribute the cost function in (1), we replace the global variable $\mathbf{s}$ which couples the per-sensor summands with auxiliary local variables $\{\mathbf{s}_j\}_{j=1}^J$ that represent candidate estimates of $\mathbf{s}$ per sensor. In conjunction with these local variables, consider the convex *constrained* minimization problem

$$\{\hat{\mathbf{s}}_j(t)\}_{j=1}^J = \arg \min_{\{\mathbf{s}_j\}_{j=1}^J} \sum_{i=1}^J E\left[(x_i(t) - \mathbf{h}_i^T(t)\mathbf{s}_i)^2\right], \quad \text{s.t. } \varepsilon_j \mathbf{s}_j = \varepsilon_j \bar{\mathbf{s}}_b, \quad b \in \mathcal{B},\ j \in \mathcal{N}_b \tag{6}$$

where $\mathcal{B} \subseteq \mathcal{J}$ is the *bridge* sensor set introduced in [18], and the additional set of consensus-enforcing variables $\{\bar{\mathbf{s}}_b\}_{b\in\mathcal{B}}$ are maintained at each of the bridge sensors comprising $\mathcal{B}$. Regarding the positive constants ε_j, though they do not cause any effect whatsoever on the constraints in (6), they will play an important role in the performance of the D-LMS algorithm (see Remark 6). Two simple conditions define a valid set $\mathcal{B}$: i) for every sensor j there exists at least one bridge sensor $b \in \mathcal{B}$ such that $b \in \mathcal{N}_j$ (the bridge neighbors of sensor j will be denoted by $\mathcal{B}_j := \mathcal{N}_j \cap \mathcal{B}$); and ii) for every two bridge sensors b_1 and b_2 there exists a path connecting them which is devoid of edges that link two nonbridge sensors. Multiple sensor assignments will qualify as valid bridge subsets for a given WSN. For instance, the set of all sensors $\mathcal{J}$ is a valid one with maximum cardinality; see also Fig. 1, where sensors in black depict $\mathcal{B}$. An upper bound on the number of bridge neighbors per sensor is provided by the maximum connectivity degree in the WSN, namely $D := \max_{j\in\mathcal{J}} |\mathcal{N}_j|$. Note that typically D is much smaller than the total number of sensors J. From a practical viewpoint, $\mathcal{B}$ can be determined and maintained in a distributed fashion using, e.g., the simple and efficient polynomial time algorithm in [25].

The WSN connectivity assumption a1) along with the defining characteristics of $\mathcal{B}$ provide necessary and sufficient conditions to assure that the equality constraints in (6) imply $\mathbf{s}_j = \mathbf{s}_{j'}\ \forall\, j, j' \in \mathcal{J}$ [18, Proposition 1]. This establishes the equivalence between (1) and (6) in the sense that their optimal solutions coincide; i.e., $\hat{\mathbf{s}}_j(t) = \hat{\mathbf{s}}(t)\ \forall\, j \in \mathcal{J}$. Two important structural properties of (6) should be appreciated, as they will be instrumental in the development of a distributed algorithm to compute $\{\hat{\mathbf{s}}_j(t)\}_{j=1}^J$: i) the separable structure of the objective function; and ii) the constraints which involve variables of neighboring sensors only.

A. Algorithm Construction

In order to solve (6), we associate Lagrange multipliers $\{\mathbf{v}_j^b\}_{j\in\mathcal{J}}^{b\in\mathcal{B}_j}$ with the corresponding equality constraints and consider the quadratically augmented Lagrangian function given by

$$\mathcal{L}_a[\boldsymbol{s}, \bar{\mathbf{s}}, \mathbf{v}] = \sum_{j=1}^J E\left[\left(x_j(t+1) - \mathbf{h}_j^T(t+1)\mathbf{s}_j\right)^2\right] + \sum_{b\in\mathcal{B}} \sum_{j\in\mathcal{N}_b} \left[\varepsilon_j \left(\mathbf{v}_j^b\right)^T (\mathbf{s}_j - \bar{\mathbf{s}}_b) + \frac{c_j \varepsilon_j^2}{2} \|\mathbf{s}_j - \bar{\mathbf{s}}_b\|^2\right] \tag{7}$$

where $\boldsymbol{s} := \{\mathbf{s}_j\}_{j=1}^J$, $\bar{\mathbf{s}} := \{\bar{\mathbf{s}}_b\}_{b\in\mathcal{B}}$, $\mathbf{v} := \{\mathbf{v}_j^b\}_{j\in\mathcal{J}}^{b\in\mathcal{B}_j}$ and $c_j > 0$ are coefficients penalizing the violation of the constraints $\varepsilon_j \mathbf{s}_j = \varepsilon_j \bar{\mathbf{s}}_b,\ \forall\, b \in \mathcal{B}$. The Lagrange multipliers $\{\mathbf{v}_j^b\}_{b\in\mathcal{B}_j}$ are maintained at sensor j. We will now resort to the AD-MoM [2, p. 253] to iteratively minimize (7) through a set of simple recursions that update $\{\boldsymbol{s}, \bar{\mathbf{s}}, \mathbf{v}\}$ in a fully distributed fashion. Because the D-LMS algorithm is designed for online estimation, the recursions will run in real-time and hence the iteration index will coincide with the time index t.

The first step consists of locally updating the Lagrange multipliers via dual gradient ascent iterations, as it is customary in the various methods of multipliers [2, Ch. 3]. The pertinent recursions are

$$\mathbf{v}_j^b(t) = \mathbf{v}_j^b(t-1) + \varepsilon_j c_j \left(\mathbf{s}_j(t) - \bar{\mathbf{s}}_b(t)\right),\ j \in \mathcal{J},\ b \in \mathcal{B}_j. \tag{8}$$

The second step involves recursions of the local estimates $\boldsymbol{s}$ obtained by minimizing (7) using block coordinate descent, i.e., $\mathcal{L}_a[\boldsymbol{s}, \bar{\mathbf{s}}, \mathbf{v}]$ is minimized with regards to $\boldsymbol{s}$ assuming all other variables $\bar{\mathbf{s}}(t) := \{\bar{\mathbf{s}}_b(t)\}_{b\in\mathcal{B}}$, and $\mathbf{v}(t) := \{\mathbf{v}_j^b(t)\}_{j\in\mathcal{J}}^{b\in\mathcal{B}_j}$ from (8) are fixed. The separable structure of (6) is inherited by the augmented Lagrangian, and therefore

$$\boldsymbol{s}(t+1) = \arg\min_{\boldsymbol{s}} \mathcal{L}_a[\boldsymbol{s}, \bar{\mathbf{s}}(t), \mathbf{v}(t)]$$

decouples into J simpler minimization subproblems

$$\mathbf{s}_j(t+1) = \arg\min_{\mathbf{s}_j} \left(E\left[\left(x_j(t+1) - \mathbf{h}_j^T(t+1)\mathbf{s}_j\right)^2\right] + \sum_{b\in\mathcal{B}_j} \left[\varepsilon_j \left(\mathbf{v}_j^b(t)\right)^T \mathbf{s}_j + \frac{c_j\varepsilon_j^2}{2} \|\mathbf{s}_j - \bar{\mathbf{s}}_b(t)\|^2\right]\right). \tag{9}$$

Since the cost in (9) is convex and differentiable, the first-order necessary condition is also sufficient for optimality. Computing the gradient with respect to $\mathbf{s}_j$ and setting the result equal to zero, yields

$$E\left[-2\mathbf{h}_j(t+1)\left(x_j(t+1) - \mathbf{h}_j^T(t+1)\mathbf{s}_j\right) + \sum_{b\in\mathcal{B}_j} \varepsilon_j \mathbf{v}_j^b(t) + \sum_{b\in\mathcal{B}_j} \varepsilon_j^2 c_j \left(\mathbf{s}_j - \bar{\mathbf{s}}_b(t)\right)\right] = \mathbf{0}. \tag{10}$$

Thus, the local estimate update $\mathbf{s}_j(t+1)$ can be obtained as the root of an equation of the form $\mathbf{f}(\mathbf{s}_j) := E[\boldsymbol{\varphi}(\mathbf{s}_j, x_j(t+1), \mathbf{h}_j(t+1))] = \mathbf{0}$, where $\boldsymbol{\varphi}(\cdot)$ stands for the function inside the expectation in (10). In lieu of local (cross-) covariance information, namely $\mathbf{r}_{h_j x_j} := E[\mathbf{h}_j(t+1)x_j(t+1)]$ and $\mathbf{R}_{h_j} := E\left[\mathbf{h}_j(t+1)\mathbf{h}_j^T(t+1)\right]$, the root of $\mathbf{f}(\mathbf{s}_j) = \mathbf{0}$ is not computable in closed form since the function $\mathbf{f}(\mathbf{s}_j)$ is unknown. Hence, motivated by stochastic approximation techniques (such as the Robbins–Monro algorithm [7, Ch. 1]), which find the root of an unknown function $\mathbf{f}(\mathbf{s}_j)$ given a time-series of noisy observations $\{\boldsymbol{\varphi}(\mathbf{s}_j(t), x_j(t+1), \mathbf{h}_j(t+1))\}_{t=0}^\infty$, the proposed recursion for all $j = 1, \ldots, J$ is

$$\mathbf{s}_j(t+1) = \mathbf{s}_j(t) + \mu_j \left[2\mathbf{h}_j(t+1)e_j(t+1) - \varepsilon_j^2 c_j |\mathcal{B}_j| \mathbf{s}_j(t) - \sum_{b\in\mathcal{B}_j} \left(\varepsilon_j \mathbf{v}_j^b(t) - \varepsilon_j^2 c_j \bar{\mathbf{s}}_b(t)\right)\right] \tag{11}$$

where μ_j is a constant step-size and $e_j(t+1) := x_j(t+1) - \mathbf{h}_j^T(t+1)\mathbf{s}_j(t)$ is the local *a priori* error.

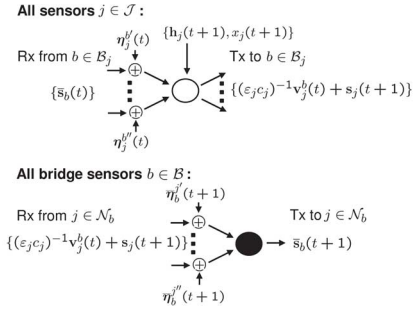

Fig. 2. D-LMS communications over noisy links.

The final step entails updating the consensus-imposing variables $\bar{\mathbf{s}}_b$ kept at the bridge sensors. The corresponding recursions are obtained by minimizing (7) with $\boldsymbol{s}(t+1) := \{\mathbf{s}_j(t+1)\}_{j\in\mathcal{J}}$ and $\mathbf{v}(t) := \left\{\mathbf{v}_j^b(t)\right\}_{j\in\mathcal{J}}^{b\in\mathcal{B}_j}$ fixed. The separability of the Lagrangian is crucial again as the general problem

$$\bar{\mathbf{s}}(t+1) = \arg\min_{\bar{\mathbf{s}}} \mathcal{L}_a[\boldsymbol{s}(t+1), \bar{\mathbf{s}}, \mathbf{v}(t)]$$

separates into $|\mathcal{B}|$ convex and differentiable equivalent subproblems

$$\bar{\mathbf{s}}_b(t+1) = \arg\min_{\bar{\mathbf{s}}_b} \sum_{j\in\mathcal{N}_b} \left[-\varepsilon_j \left(\mathbf{v}_j^b(t)\right)^T \bar{\mathbf{s}}_b + \frac{c_j\varepsilon_j^2}{2}\|\mathbf{s}_j(t+1) - \bar{\mathbf{s}}_b\|^2 \right]. \quad (12)$$

It should be noted that the expectation term in (7) has been discarded in the process of obtaining (12), since it is not dependent on $\bar{\mathbf{s}}_b$ and thus inconsequential for the minimization. Applying the first-order optimality condition explicitly yields

$$\bar{\mathbf{s}}_b(t+1) = \left(\sum_{r\in\mathcal{N}_b} \varepsilon_r^2 c_r\right)^{-1} \sum_{j\in\mathcal{N}_b} \left(\varepsilon_j \mathbf{v}_j^b(t) + \varepsilon_j^2 c_j \mathbf{s}_j(t+1)\right) \quad (13)$$

for $b \in \mathcal{B}$. Recursions (8), (11), and (13) constitute the D-LMS algorithm, which can be arbitrarily initialized. At the beginning of the tth iteration, sensor j receives the consensus variables $\bar{\mathbf{s}}_b(t)$ from its bridge neighbors $b \in \mathcal{B}_j$. With this information and using (8), it is able to update its Lagrange multipliers $\{\mathbf{v}_j^b(t)\}_{b\in\mathcal{B}_j}$ which are then jointly used along with the newly acquired local data $\{x_j(t+1), \mathbf{h}_j(t+1)\}$ to compute $\mathbf{s}_j(t+1)$ via (11). Then sensor j transmits the vector $(\varepsilon_j c_j)^{-1}\mathbf{v}_j^b(t) + \mathbf{s}_j(t+1)$ to all bridge sensors in its neighborhood $\mathcal{B}_j$. Subsequently, each sensor $b \in \mathcal{B}$ receives the vectors $(\varepsilon_j c_j)^{-1}\mathbf{v}_j^b(t) + \mathbf{s}_j(t+1)$ and scales them with $\varepsilon_j^2 c_j$ in order to find the weighted average in (13) and obtain $\bar{\mathbf{s}}_b(t+1)$, thus completing the tth iteration. Further, observe that in order to compute the weights in (13), bridge sensor b should acquire $\left\{\varepsilon_j^2 c_j\right\}_{j\in\mathcal{N}_b}$ only from its neighbors during the startup phase of the WSN.

Communications take place among single-hop neighboring sensors only, at a resulting cost that scales linearly in p, the dimensionality of $\mathbf{s}_0$. Incorporating also the effects of additive communication noise, Fig. 2 depicts the vector exchanges required by D-LMS on a per iteration basis, and explicitly shows the additional tasks performed by the sensors in $\mathcal{B}$. The modified D-LMS recursions accounting for the noise corrupted variables exchanged among sensors are summarized below, and tabulated as Algorithm 1. For all sensors $j \in \mathcal{J}$ and $b \in \mathcal{B}$, the D-LMS algorithm in the noisy setup becomes

$$\mathbf{v}_j^b(t) = \mathbf{v}_j^b(t-1) + \varepsilon_j c_j \left(\mathbf{s}_j(t) - (\bar{\mathbf{s}}_b(t) + \boldsymbol{\eta}_j^b(t))\right) \quad (14)$$

$$\mathbf{s}_j(t+1) = \mathbf{s}_j(t) + \mu_j \Bigg[2\mathbf{h}_j(t+1)e_j(t+1) - \varepsilon_j^2 c_j |\mathcal{B}_j| \mathbf{s}_j(t) - \sum_{b\in\mathcal{B}_j} \left(\varepsilon_j \mathbf{v}_j^b(t) - \varepsilon_j^2 c_j (\bar{\mathbf{s}}_b(t) + \boldsymbol{\eta}_j^b(t))\right) \Bigg] \quad (15)$$

$$\bar{\mathbf{s}}_b(t+1) = \left(\sum_{r\in\mathcal{N}_b} \varepsilon_r^2 c_r\right)^{-1} \sum_{j\in\mathcal{N}_b} \Big(\varepsilon_j \mathbf{v}_j^b(t) + \varepsilon_j^2 c_j (\mathbf{s}_j(t+1) + \bar{\boldsymbol{\eta}}_b^j(t+1))\Big). \quad (16)$$

D-LMS entails $|\mathcal{B}_j|$ more recursions per sensor when compared to diffusion LMS in [9]. However, since $|\mathcal{B}_j|$ is typically much smaller than J the increase in computational complexity is relatively low. On the other hand, this additional cost and introduced hierarchy among sensors pays off with improved convergence rates as will become apparent in the numerical examples of Section III-C.

Algorithm 1: D-LMS

Arbitrarily initialize $\{\mathbf{s}_j(0)\}_{j=1}^J$, $\{\bar{\mathbf{s}}_b(0)\}_{b\in\mathcal{B}}$, and $\left\{\mathbf{v}_j^b(-1)\right\}_{j\in\mathcal{J}}^{b\in\mathcal{B}_j}$.

for $t = 0, 1, \ldots$ **do**

Bridge sensors $b \in \mathcal{B}$: transmit $\bar{\mathbf{s}}_b(t)$ to neighbors in $\mathcal{N}_b$.

All $j \in \mathcal{J}$: update $\left\{\mathbf{v}_j^b(t)\right\}_{b\in\mathcal{B}_j}$ using (14).

All $j \in \mathcal{J}$: update $\mathbf{s}_j(t+1)$ using (15).

All $j \in \mathcal{J}$: transmit $(\varepsilon_j c_j)^{-1}\mathbf{v}_j^b(t) + \mathbf{s}_j(t+1)$ to each $b \in \mathcal{B}_j$.

Bridge sensors $b \in \mathcal{B}$: compute $\bar{\mathbf{s}}_b(t+1)$ using (16).

end for

Remark 3 (Comparison With [17] and [18]): In contrast to the D-MLE, D-BLUE, and D-LMMSE schemes in [17] and [18], it is apparent that D-LMS in (14)–(16) allows online incorporation and processing of sensor data. It is an adaptive estimation/tracking algorithm, whereas the distributed estimation schemes in [17] and [18] operate in batch mode. Further, note that in D-LMS the requirement for statistical information is bypassed in the stochastic approximation step where the process statistics are learnt "on-the-fly." This is not the case in [17] and [18] where all proposed schemes are applicable as long as data models are available across sensors.

Remark 4 (Versatility Through the Use of Bridge Sensors): The bridge sensor set provides flexibility to trade-off communication cost for robustness to sensor failures. In D-LMS each sensor transmits $|\mathcal{B}_j|p$ scalars whereas in diffusion LMS each sensor transmits p scalars. However, note that in D-LMS the nonbridge sensors have to communicate (transmit and receive) with approximately half of their neighbors, namely those in $\mathcal{B}_j$. This follows since the two conditions defining the set $\mathcal{B}$ are satisfied when $|\mathcal{B}| \approx J/2$. Intuitively, this holds because if a sensor is designated to serve as a bridge then its neighbors do not have to be in $\mathcal{B}$, whereas if a sensor does not have bridge neighbors then it turns itself into one; see also the numerical tests in

[25]. With reference to Fig. 2, a nonbridge sensor j has to remain active over $2|\mathcal{B}_j|$ time slots in order to send and receive information from $b \in \mathcal{B}_j$. Among these slots, $|\mathcal{B}_j|$ are required to transmit $(\varepsilon_j c_j)^{-1}\mathbf{v}_j^b(t)+\mathbf{s}_j(t+1)$ to its bridge neighbors, while the remaining $|\mathcal{B}_j|$ to receive $\bar{\mathbf{s}}_b(t)$ from them. Bridge sensor b remains active over $2(|\mathcal{B}_b|-1)$ time slots, as any other sensor, plus $|\mathcal{N}_b|$ additional time slots: one to transmit $\bar{\mathbf{s}}_b(t)$ to its neighbors in $\mathcal{N}_b$, and the rest to receive $(\varepsilon_j c_j)^{-1}\mathbf{v}_j^b(t)+\mathbf{s}_j(t+1)$ from them. Assuming equal battery capacities across sensors, bridge sensors are expected to fail first. In diffusion LMS, each sensor has to be active over $|\mathcal{N}_j|$ time slots among which one slot is spent for transmission and the rest for listening. Now, consider the WSN in Fig. 1 where bridge sensors are disconnected. Thus, $|\mathcal{B}_b| = 1$ and the communication cost of a bridge sensor is also $|\mathcal{N}_b|$. Because typically $2|\mathcal{B}_j| \leq |\mathcal{N}_j|$, when diffusion LMS is applied to the same WSN the total number of required active time slots will be larger (31 versus 26 in this example). Thus, utilization of bridge sensors offers the potential of increasing the life expectancy of the network.[1]

Regarding recovery from sensor failures, D-LMS remains operational so long as each sensor adjusts its local recursions (14)–(16) to the modified neighborhood structure, and the overall network graph remains connected. In a possible bridge sensor failure, it might be the case that some sensors need to be promoted to $\mathcal{B}$ using the algorithm in, e.g., [25]. Thus, the network as an autonomous entity is capable of adapting to changes in the topology. The steps of the simple recovery process are given in [17, Remark 2].

Remark 5 (Consensus and Communication Protocols): Similar to all consensus-based schemes, D-LMS requires an underlying communication protocol that controls information exchanges among sensors. One feasible choice (not necessarily the most efficient) could be a time division multiple-access (TDMA) system, where each sensor is allocated a time slot during which it can transmit data to its neighbors that operate in reception mode. Consider a TDMA system with $J+|\mathcal{B}|$ time slots. During the first $|\mathcal{B}|$ slots each bridge sensor b transmits to all its neighbors its consensus variable $\bar{\mathbf{s}}_b(t)$ required in (14) and (15). Each of the p scalars in $\bar{\mathbf{s}}_b(t)$ can be transmitted using, e.g., multicarrier modulation. Then, during the $(|\mathcal{B}|+j)$th time slot only sensor j is active and broadcasts $(\varepsilon_j c_j)^{-1}\mathbf{v}_j^b(t)+\mathbf{s}_j(t+1)$ to each of its bridge neighbors. Recalling that $|\mathcal{B}_j| \ll J$, sensor j could either use different frequency bands to transmit information to each of its bridge neighbors, or, could devote $(1/|\mathcal{B}_j|)$th fraction of the time slot for each of its $|\mathcal{B}_j|$ bridge neighbors. Such a scheme requires i) unique sensor indexing established prior to the WSN deployment and ii) global synchronization across the WSN, for which there are available algorithms in the existing literature, e.g., see [24].

B. Consensus Controller Interpretation

Even though recursions (14)–(16) clearly suggest simplicity as an asset of the proposed algorithm, they may somehow obscure the essential mechanisms operating on the available information to yield the estimates $\mathbf{s}_j$. Here we derive a set of equivalent recursions which turn out to be insightful about these issues, despite being less appropriate for online implementation than (14)–(16).

For arbitrary $j \in \mathcal{J}$ and $b \in \mathcal{B}_j$, consider the noise-free Lagrange multiplier update recursion (8) with initial condition $\mathbf{v}_j^b(-1) = \mathbf{0}$. By recognizing $\mathbf{v}_j^b(t)$ as the output of an accumulator system whose input $\varepsilon_j c_j\,[\mathbf{s}_j(t) - \bar{\mathbf{s}}_b(t)]$ is the sequence of scaled constraint violations, the zero initial condition yields the equivalent nonrecursive form [cf. (8)]

$$\mathbf{v}_j^b(t) = \sum_{n=0}^{t} \varepsilon_j c_j\,[\mathbf{s}_j(n) - \bar{\mathbf{s}}_b(n)]\,. \tag{17}$$

Arguing by induction as in [18, Lemma 3], the consensus variables for all $b \in \mathcal{B}$ and $t \geq 0$ can be expressed as [cf. (13)]

$$\bar{\mathbf{s}}_b(t) = \left(\sum_{r\in\mathcal{N}_b} \varepsilon_r^2 c_r\right)^{-1} \sum_{j\in\mathcal{N}_b} \varepsilon_j^2 c_j \mathbf{s}_j(t). \tag{18}$$

Equation (18) establishes that the consensus variables $\bar{\mathbf{s}}_b(t)$ are simply obtained as a weighted average of the local estimates gathered from sensor b's neighborhood.

Consider now the vector $\mathbf{q}_j(t) := \mathbf{s}_j(t) - |\mathcal{B}_j|^{-1}\sum_{b\in\mathcal{B}_j}\bar{\mathbf{s}}_b(t)$, which represents the *instantaneous consensus error* at sensor j, as measured with respect to the *consensus reference* given by the average $|\mathcal{B}_j|^{-1}\sum_{b\in\mathcal{B}_j}\bar{\mathbf{s}}_b(t)$. Setting the penalty coefficients as $c_j = 1/|\mathcal{B}_j|$ and using (17) to eliminate the Lagrange multipliers from (11) yields

$$\mathbf{s}_j(t+1) = \mathbf{s}_j(t) + \mu_j 2\mathbf{h}_j(t+1)e_j(t+1) - \mu_j\varepsilon_j^2\mathbf{q}_j(t) - \mu_j\varepsilon_j^2\sum_{n=0}^{t}\mathbf{q}_j(n). \tag{19}$$

Equations (18) and (19) are equivalent to D-LMS under ideal links, when $\mathbf{v}_j^b(-1) = \mathbf{0}$. As they stand, the new recursions are not suitable for real-time implementation because the sum term in (19) requires storing the entire history of $\mathbf{q}_j(t)$. Nonetheless, they shed light into the signal processing taking place at each sensor, which turns out to be remarkably intuitive as discussed next.

The right-hand side (rhs) of (19) readily suggests that the local estimate $\mathbf{s}_j(t+1)$ is obtained as the superposition of three terms: a) the sum $\mathbf{s}_j(t) + \mu_j 2\mathbf{h}_j(t+1)e_j(t+1)$ represents a *local* LMS adaptation based on the new information $\{\mathbf{h}_j(t+1), x_j(t+1)\}$ available at sensor j; b) an update based on a proportional correction $\mu_j\varepsilon_j^2\mathbf{q}_j(t)$ due to the instantaneous consensus error $\mathbf{q}_j(t)$; and c) a correction sum due to the accumulated consensus error (discrete-time integral). A term like a) is expected, whereas the rest should explain the mechanisms employed to incorporate the extra information gathered from the whole WSN. In fact, b) and c) show that a proportional-integral (PI) discrete-time controller, see, e.g., [3, p. 605], is used to drive the local estimate $\mathbf{s}_j(t)$ to consensus, as dictated by the computed time-varying set-point $|\mathcal{B}_j|^{-1}\sum_{b\in\mathcal{B}_j}\bar{\mathbf{s}}_b(t)$; see also Fig. 3. It is exclusively throughout this reference that global information is percolated to improve the local estimate $\mathbf{s}_j$.

The first closed-loop system interpretation for consensus schemes was given in [5]. Let $\tilde{\mathbf{q}}_j(t) := |\mathcal{B}_j|\mathbf{q}_j(t) = \sum_{b\in\mathcal{B}_j}[\mathbf{s}_j(t) - \bar{\mathbf{s}}_b(t)]$ and eliminate $\bar{\mathbf{s}}_b(t)$ using (18). This leads to the global representation $\tilde{\mathbf{q}}(t) = \mathbf{A}\mathbf{s}(t)$ with

[1]We reiterate, however, that the results in this paper carry over even when every sensor acts also as bridge sensor, i.e., $\mathcal{B} = \mathcal{J}$.

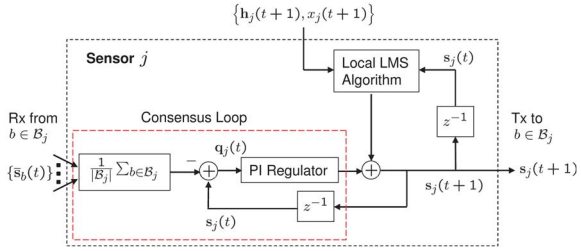

Fig. 3. D-LMS sensor processing to obtain local estamates $\mathbf{s}_j$.

$\tilde{\mathbf{q}}(t) := [\tilde{\mathbf{q}}_1^T(t)\dots\tilde{\mathbf{q}}_J^T(t)]^T$, $\mathbf{s}(t) := [\mathbf{s}_1^T(t)\dots\mathbf{s}_J^T(t)]^T$, and the generalized "two-hop range" Laplacian is given by

$$\mathbf{A} := \mathrm{bdiag}(|\mathcal{B}_1|\mathbf{I}_p,\dots,|\mathcal{B}_J|\mathbf{I}_p) - \sum_{b\in\mathcal{B}} \frac{(\mathbf{e}_b\otimes\mathbf{I}_p)(\mathbf{e}_b\otimes\mathbf{I}_p)^T}{\sum_{r\in\mathcal{N}_b}\varepsilon_r^2 c_r}\mathrm{bdiag}\left(\varepsilon_1^2 c_1\mathbf{I}_p,\dots,\varepsilon_J^2 c_J\mathbf{I}_p\right) \tag{20}$$

where $\mathbf{e}_b$ represents the bth column of the adjacency matrix $\mathbf{E}$. Indeed, under a1) $\mathbf{A}$ shares fundamental properties with an undirected graph Laplacian, i.e., it is symmetric positive semidefinite and its null space is the consensus subspace (vectors with equal entries). The network-wide feedback $-\mu\mathbf{A}\mathbf{s}(t)$ [cf. (19) and Fig. 3] links D-LMS with the standard Laplacian-based consensus protocol studied in e.g., [5] and [12], whereas the difference stems from the use of bridge sensors. This may not be surprising if one recalls that the Laplacian-based protocol for undirected graphs can be derived using an iterative procedure minimizing a suitable disagreement potential [12, eq. (5)]. The latter strongly resembles the quadratic term augmenting the Lagrangian in (7). The extended two-hop information range enjoyed by D-LMS should be also contrasted with diffusion LMS [9], which only spans the single-hop neighborhood. Though, as clarified in Remark 4 a small price is paid in terms of the amount of data needed to be transmitted from each sensor.

Remark 6 (Consensus Loop Tuning): The constant ε_j is only affecting the PI gains of the consensus regulator [cf. (19)]. For $\varepsilon_j = 1$ these gains boil down to μ_j, a generally small constant attenuating the influence that the information embedded in b) and c) has on the estimate $\mathbf{s}_j(t+1)$. The presence of ε_j is thus intuitively justified as a compensator for this effect, gaining an additional degree of freedom to attain potentially faster convergence and/or better estimation performance. Indeed, our simulation results in [10] corroborate considerable improvements when selecting $\varepsilon_j^2\mu_j \approx 1$. For a given step-size and contrasting with $\varepsilon_j = 1$, the steady-state estimation error is markedly reduced at a modest price slightly decreasing the convergence rate of the MSE cost in (1). On the other hand, if the D-LMS step-size is increased to the point that there is no gain in estimation error, then the MSE reaches steady-state much faster without a noticeable misadjustment. Based on a suitable performance criterion, a problem falling outside the scope of this paper, it would be interesting to optimally design the ε_j coefficients; see [26] for a related weight optimization approach in the context of consensus averaging problems.

C. Numerical Examples

Here, we test the novel D-LMS algorithm, and compare its global MSE performance with i) diffusion LMS using Metropolis weights [9]; ii) local (L-) LMS whereby each sensor runs an independent LMS filter using its local information only (no communications); iii) centralized incremental LMS [8]; and iv) C-LMS [cf. (2)]. The WSN is simulated as a $\mathcal{G}^2(80, 0.6)$ graph, and for the examples with noisy links receiver AWGN with variance $\sigma_\eta^2 = 10^{-3}$ is added. The signal vector $\mathbf{s}_0 = \mathbf{1}_p$ has dimensionality $p = 8$, and for all $j \in \mathcal{J}$ the regressors $\mathbf{h}_j(t) = [h_j(t)\dots h_j(t-p+1)]^T$ have entries that evolve according to $h_j(t) = (1-\rho)u_{1,j}h_j(t-1) + \sqrt{\rho}\nu_j(t)$. We choose $\rho = 2\times 10^{-1}$, the $u_{1,j} \sim \mathcal{U}[0,1]$ (uniformly distributed) are i.i.d. in space, and the driving white noise $\nu_j(t) \sim \mathcal{U}\left[-\sqrt{3}\sigma_{\nu_j}, \sqrt{3}\sigma_{\nu_j}\right]$ has a spatial variance profile given by $\sigma_{\nu_j}^2 = 10^{-1}u_{2,j}$ with $u_{2,j} \sim \mathcal{U}[0,1]$ and i.i.d. A linear model $\mathbf{x}(t) = \mathbf{H}(t)\mathbf{s}_0 + \boldsymbol{\epsilon}(t)$ is adopted with observation WGN of spatial variance profile $\sigma_{\epsilon_j}^2 = 10^{-4}u_{3,j}$, with i.i.d. $u_{3,j} \sim \mathcal{U}[0,1]$. For all four algorithms the step-size is set to $\mu = 9\times 10^{-2}$, and in particular for D-LMS $c_j = \varepsilon_j = 1\ \forall\ j \in \mathcal{J}$.

Fig. 4 (top) compares the normalized MSE evolution (learning curve) obtained as $J^{-1}\sum_{j=1}^J E[e_j(t)^2]$ for the distributed schemes, where the expectation is approximated by averaging 50 Monte Carlo realizations. Both incremental LMS and C-LMS provide a comparable performance benchmark while L-LMS stands on the other extreme. For both distributed approaches in the (communication) noise-free setting, the resultant misadjustment is negligible thus matching the performance (in this sense) of its centralized counterparts. Furthermore, D-LMS outperforms diffusion LMS whereas its MSE remains bounded even when channel links are corrupted by reception noise, with an inflated steady-state MSE level, as expected.

To gauge local performance, we evaluate the figures of merit which are customary in the adaptive literature [8], [9]: i) MSE $E[e_j(t)^2]$, ii) excess-MSE (EMSE) $E[e_j(t)^2] - \sigma_{\epsilon_j}^2$, and iii) mean-square deviation (MSD) $E[\|\mathbf{s}_j(t) - \mathbf{s}_0\|^2]$. For the previous setup, the steady-state values of these metrics are depicted in Fig. 4 (bottom). Good performance for D-LMS is apparent from the MSE curve which comes very close to the local noise levels; observe also the small EMSE. Sensor collaboration, despite the diverse noise statistical profile across the WSN, smoothens network-wide EMSE/MSD values. In comparison with diffusion LMS, D-LMS exhibits a slight edge on MSD while EMSE levels are comparable.

Under the same WSN setup, we illustrate the capabilities of D-LMS when it comes to tracking a time-varying signal vector $\mathbf{s}_0(t)$. The large amplitude slowly time-varying process model $\mathbf{s}_0(t) = (1-\rho)\mathbf{s}_0(t-1) + \sqrt{\rho}\boldsymbol{\nu}(t)$ is simulated, with $\rho = 10^{-2}$ and $\boldsymbol{\nu}(t) \sim \mathcal{N}(\mathbf{0}, 2\times 10^{-2}\mathbf{I}_8)$. Fig. 5 (top) depicts the fifth and third entries of the true time-varying parameter $\mathbf{s}_0(t)$, and the respective estimates from sensors 38 and 70 that closely follow the true variations. Both sensors and parameter entries were chosen uniformly at random, in the interest of showing the representative behavior across the WSN. In addition, we also plot the estimates $[\mathbf{s}_{38}(t)]_5, [\mathbf{s}_{70}(t)]_3$ obtained when the WSN model accounts for communication noise with $\sigma_\eta^2 = 10^{-2}$. Larger estimate fluctuations are a direct manifestation of the increased MSE.

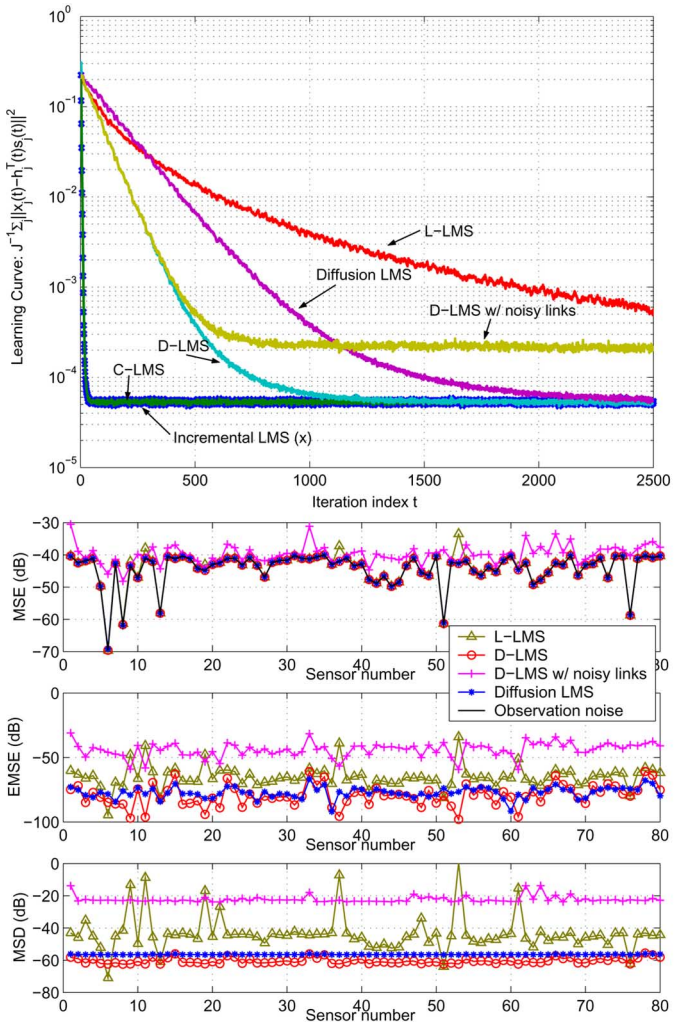

Fig. 4. (top) Normalized global MSE (learning curve). (bottom) Local performance figures of merit: MSE, EMSE, and MSD.

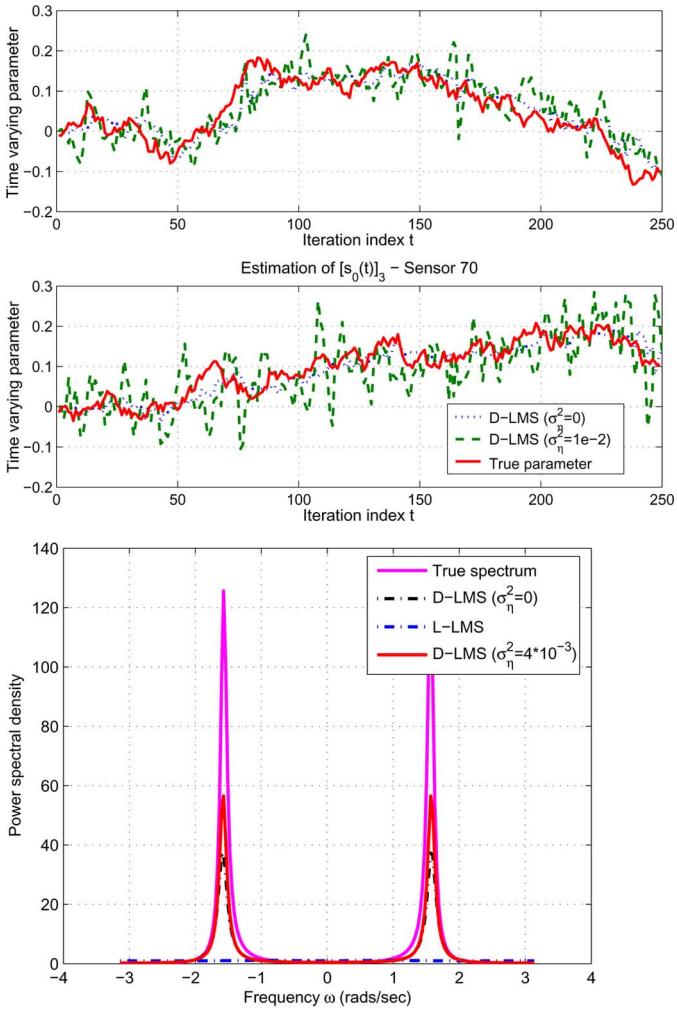

Fig. 5. (top) Tracking with D-LMS. (bottom) Spectral estimation with D-LMS.

Next, we examine the D-LMS performance in the spectrum estimation application described earlier in Remark 1 for the aforementioned WSN setting. The AR source has order $p = 4$ and coefficients $\mathbf{s}_0 = [-0.31, 1.14, 0.275, 0.222]^T$. The source signal propagates via multipath channels of order $L_j = 2$ and arrives at the sensors where it gets contaminated with sensing noise having variance 10^{-4} [cf. (4)]. The channels corresponding to sensors 3, 7, 15, 27, 37, 57, 67 are set so that they have a null at the frequency where the AR source has a peak, namely at $\omega = \pi/2$. Fig. 5 (bottom) depicts the actual power spectral density of the source as well as the estimated ones at sensor 15 using L-LMS and D-LMS under ideal and noisy intersensor links. The step-size is $\mu = 0.98 \times 10^{-3}$, while $\varepsilon_j = c_j = 1$. Clearly, even in the presence of communication noise D-LMS exploits the spatial diversity available and allows all sensors to estimate accurately the actual spectral peak, whereas L-LMS leads the problematic sensors, e.g., sensor 15 in Fig. 5 (bottom), to misleading estimates. The latter corroborates the ability of D-LMS to percolate information across the WSN.

IV. Stability and Performance Analysis

An attractive feature of D-LMS is that it can be applied to a wide class of signals. Indeed, D-LMS requires no assumption on the statistics of $\{x_j(t), \mathbf{h}_j(t)\}_{j=1}^J$. When it comes to stability and performance evaluation however, a meaningful "ground-truth" model should be adopted to carry out the analysis and enable fair comparison among competing alternatives. Toward this end, we adopt the standard data model, commonly used throughout the adaptive signal processing literature, e.g., [8], [9], [20], and [21, Ch. 5,9]:

a2) *The sensor observations adhere to the linear model*

$$x_j(t) = \mathbf{h}_j^T(t)\mathbf{s}_0 + \epsilon_j(t), \quad j \in \mathcal{J} \tag{21}$$

where the white noise $\epsilon_j(t)$ is zero-mean with variance $\sigma_{\epsilon_j}^2$.

In order to facilitate stability analysis, an important preliminary step is to express D-LMS as a linear time-varying (LTV) stochastic difference-equation. Specifically, starting from (14)–(16) and applying simple algebraic manipulations we will obtain recursions for the local estimation errors $\mathbf{y}_{1,j}(t) := \mathbf{s}_j(t) - \mathbf{s}_0$ and the local sum of multipliers $\mathbf{y}_{2,j}(t) := \sum_{b \in \mathcal{B}_j} \mathbf{v}_j^b(t-1)\, \forall\, j \in \mathcal{J}$. For simplicity in exposition, set $\varepsilon_j^2 c_j = \mu_j^{-\gamma} c_1/2$, $\varepsilon_j c_j = \mu_j^{-\gamma/2} c_2$ for any $\gamma \in (0,1)$, and $\mu_j = \mu\, \forall\, j \in \mathcal{J}$. Such a selection of ε_j is well motivated, since it gives more emphasis to the information gathered from the neighborhood [cf. (19)]. This is desirable in WSN-based

applications including the one described in Remark 1, where sensor collaboration is essential to efficiently estimate the parameters of interest. Further, stack $\{\mathbf{y}_{1,j}(t), \mathbf{y}_{2,j}(t)\}_{j=1}^{J}$ vectors to form the $Jp \times 1$ supervectors $\mathbf{y}_1(t) := \left[\mathbf{y}_{1,1}^T(t) \ldots \mathbf{y}_{1,J}^T(t)\right]^T$ and $\mathbf{y}_2(t) := \left[\mathbf{y}_{2,1}^T(t) \ldots \mathbf{y}_{2,J}^T(t)\right]^T$, respectively; and let $\mathbf{y}(t) := \left[\mathbf{y}_1^T(t) \;\; \mathbf{y}_2^T(t)\right]^T \in \mathbb{R}^{2Jp \times 1}$. The Lagrange multipliers $\mathbf{v}_j^b(-1)$ are initialized so that

$$\sum_{j \in \mathcal{N}_b} \mathbf{v}_j^b(-1) = \mathbf{0}, \quad \forall\, b \in \mathcal{B}. \tag{22}$$

Equation (22) is easy to satisfy since sensors can initialize arbitrarily their multipliers, e.g., through zero initial conditions whereby sensor j sets $\mathbf{v}_j^b(-1) = \mathbf{0}$ with $b \in \mathcal{B}_j$.

The next step is to rewrite the consensus variable recursion in (16) in a way that will allow us later on to derive a first-order recursion for $\mathbf{y}(t)$. Specifically, we prove in Appendix A the following lemma.

Lemma 1: *If the Lagrange multipliers $\mathbf{v}_j^b(t)$ are initialized as in (22), then the consensus variables $\bar{\mathbf{s}}_b(t)$ can be expressed for $t \geq -1$ as*

$$\bar{\mathbf{s}}_b(t+1) = \frac{1}{|\mathcal{N}_b|} \sum_{j \in \mathcal{N}_b} \mathbf{s}_j(t+1) + \frac{1}{|\mathcal{N}_b|} \sum_{j \in \mathcal{N}_b} \left[\bar{\boldsymbol{\eta}}_b^j(t+1) - \bar{\boldsymbol{\eta}}_b^j(t)\right] - \frac{1}{|\mathcal{N}_b|} \sum_{j \in \mathcal{N}_b} \boldsymbol{\eta}_j^b(t), \quad b \in \mathcal{B} \tag{23}$$

with $\boldsymbol{\eta}_j^b(-1) = \bar{\boldsymbol{\eta}}_b^j(-1) = \mathbf{0}$, and $\bar{\boldsymbol{\eta}}_b^b(t) = \boldsymbol{\eta}_b^b(t) = \mathbf{0}$.

Using Lemma 1, as well as the multiplier update rule in (14) we wish to derive first-order recursions for $\mathbf{y}_{1,j}(t) := \mathbf{s}_j(t) - \mathbf{s}_0$ and $\mathbf{y}_{2,j}(t) := \sum_{b \in \mathcal{B}_j} \mathbf{v}_j^b(t-1)$. To this end, let us define the noise vectors

$$\bar{\boldsymbol{\eta}}_j(t) := \sum_{b \in \mathcal{B}_j} \sum_{j' \in \mathcal{N}_b} \frac{\bar{\boldsymbol{\eta}}_b^{j'}(t)}{|\mathcal{N}_b|}, \quad j \in \mathcal{J} \tag{24}$$

that depend on the reception noise at the bridge neighbors. Further, consider two more noise vectors

$$\boldsymbol{\eta}_j^{\alpha}(t) := \sum_{b \in \mathcal{B}_j} \boldsymbol{\eta}_j^b(t), \quad \boldsymbol{\eta}_j^{\beta}(t) := \sum_{b \in \mathcal{B}_j} \sum_{j' \in \mathcal{N}_b} \frac{\boldsymbol{\eta}_{j'}^b(t)}{|\mathcal{N}_b|}. \tag{25}$$

Based on definitions (24) and (25), it is shown in Appendix B that:

Lemma 2: *Under a2) and with $\left\{\mathbf{v}_j^b(-1)\right\}_{j \in \mathcal{N}_b}$ selected to satisfy (22), local state vectors $\{\mathbf{y}_{i,j}(t)\}_{j=1}^{J}$ obey the recursions*

$$\begin{aligned}\mathbf{y}_{1,j}(t+1) = {} & \left(\mathbf{I}_p - 2\mu \mathbf{h}_j(t+1)\mathbf{h}_j^T(t+1) - \mu^{1-\gamma} c_1 |\mathcal{B}_j| \mathbf{I}_p\right)\mathbf{y}_{1,j}(t) \\ & + \mu^{1-\gamma} c_1 \left[\sum_{b \in \mathcal{B}_j} \left(\frac{1}{|\mathcal{N}_b|} \sum_{j' \in \mathcal{N}_b} \mathbf{y}_{1,j'}(t)\right) - \frac{\mu^{\gamma/2}}{2c_2} \mathbf{y}_{2,j}(t) \right. \\ & \left. + \left(\bar{\boldsymbol{\eta}}_j(t) - \bar{\boldsymbol{\eta}}_j(t-1)\right) + \left(\boldsymbol{\eta}_j^{\alpha}(t) - \boldsymbol{\eta}_j^{\beta}(t-1)\right)\right] \\ & + 2\mu \mathbf{h}_j(t+1)\epsilon_j(t+1)\end{aligned} \tag{26}$$

$$\begin{aligned}\mathbf{y}_{2,j}(t+1) = {} & \mathbf{y}_{2,j}(t) + \mu^{-\gamma/2} c_2 \\ & \times \left[|\mathcal{B}_j| \mathbf{y}_{1,j}(t) - \sum_{b \in \mathcal{B}_j} \left(\frac{1}{|\mathcal{N}_b|} \sum_{j' \in \mathcal{N}_b} \mathbf{y}_{1,j'}(t)\right) \right. \\ & \left. - \left(\bar{\boldsymbol{\eta}}_j(t) - \bar{\boldsymbol{\eta}}_j(t-1)\right) - \left(\boldsymbol{\eta}_j^{\alpha}(t) - \boldsymbol{\eta}_j^{\beta}(t-1)\right)\right].\end{aligned} \tag{27}$$

Aiming at a first-order recursion for $\mathbf{y}(t)$, consider concatenating the noise terms in (24) and (25) for $j = 1, \ldots, J$ to form the $Jp \times 1$ supervectors $\bar{\boldsymbol{\eta}}(t)$, $\boldsymbol{\eta}^{\alpha}(t)$ and $\boldsymbol{\eta}^{\beta}(t)$, respectively; and also define the global observation noise vector

$$\boldsymbol{\epsilon}(t) := 2\mu[\mathbf{h}_1^T(t)\epsilon_1(t) \ldots \mathbf{h}_J^T(t)\epsilon_J(t)]^T. \tag{28}$$

Upon stacking $\mathbf{y}_{1,j}(t+1)$ and $\mathbf{y}_{2,j}(t+1)$ from Lemma 2, for $j = 1, \ldots, J$, in $\mathbf{y}(t+1)$, it is shown in Appendix C that the D-LMS recursions (14)–(16) can be compactly written in matrix form as

$$\begin{aligned}\mathbf{y}(t+1) = {} & \boldsymbol{\Psi}(t+1, \mu)\mathbf{y}(t) + \begin{bmatrix} \mu^{1-\gamma} c_1 \mathbf{I}_{Jp} & \mu^{1-\gamma} c_1 \mathbf{I}_{Jp} \\ -\mu^{\gamma/2} c_2 \mathbf{I}_{Jp} & -\mu^{\gamma/2} c_2 \mathbf{I}_{Jp} \end{bmatrix} \\ & \times \begin{bmatrix} \bar{\boldsymbol{\eta}}(t) - \bar{\boldsymbol{\eta}}(t-1) \\ \boldsymbol{\eta}^{\alpha}(t) - \boldsymbol{\eta}^{\beta}(t-1) \end{bmatrix} + \begin{bmatrix} \boldsymbol{\epsilon}(t+1) \\ \mathbf{0} \end{bmatrix}\end{aligned} \tag{29}$$

where for $t \geq 0$ the $2Jp \times 2Jp$ transition matrix $\boldsymbol{\Psi}(t, \mu)$ consists of four $Jp \times Jp$ matrix blocks given by $[\boldsymbol{\Psi}(t, \mu)]_{11} = \mathbf{I}_{Jp} - 2\mu \mathbf{R}_h(t) - \mu^{1-\gamma} c_1 \mathbf{A}$, $[\boldsymbol{\Psi}(t, \mu)]_{12} = -\left(\mu^{1-\gamma/2} c_1/(2c_2)\right)\mathbf{I}_{Jp}$, $[\boldsymbol{\Psi}(t, \mu)]_{21} = \mu^{-\gamma/2} c_2 \mathbf{A}$ and $[\boldsymbol{\Psi}(t, \mu)]_{22} = \mathbf{I}_{Jp}$, with $\mathbf{R}_h(t) := \mathrm{bdiag}\left(\mathbf{h}_1(t)\mathbf{h}_1^T(t), \ldots, \mathbf{h}_J(t)\mathbf{h}_J^T(t)\right)$ while the structure of $\mathbf{A}$ is given in (20).

The linear dynamical system described by (29) is indeed time-varying since $\boldsymbol{\Psi}(t, \mu)$ as well as $\mathbf{R}_h(t)$ are time dependent. It is also random due to the noise terms as well as the regression vectors $\mathbf{h}_j(t)$. In order to satisfy the initialization requirement in (22), $\mathbf{y}_{2,j}(0)$ should be set to $\mathbf{y}_{2,j}(0) = \sum_{b \in \mathcal{B}_j} \mathbf{v}_j^b(-1) = \mu^{-\gamma/2} c_2 \mathbf{A}_j \mathbf{y}_2'(0)$, where $\mathbf{A}_j$ is the jth $p \times Jp$ block of the decomposition $\mathbf{A} = \left[\mathbf{A}_1^T \ldots \mathbf{A}_J^T\right]^T$ and $\mathbf{y}_2'(0)$ can be chosen arbitrarily. Hence, using

$$\mathbf{y}_{2,j}(0) = \mu^{-\gamma/2} c_2 \sum_{b \in \mathcal{B}_j} \left[\mathbf{y}_{2,j}'(0) - \frac{1}{|\mathcal{N}_b|} \sum_{j' \in \mathcal{N}_b} \mathbf{y}_{2,j'}'(0)\right] \tag{30}$$

and setting $\mathbf{v}_j^b(-1) = \mathbf{y}_{2,j}'(0) - |\mathcal{N}_b|^{-1} \sum_{j' \in \mathcal{N}_b} \mathbf{y}_{2,j'}'(0)$ ensures that (22) holds true, while $\mu^{-\gamma/2} c_2$ is placed for normalization.

The LTV system in (29) is not yet ready for stability analysis since $E_{\mathbf{h}}[\boldsymbol{\Psi}(t, \mu)]$ does not have all its eigenvalues inside the unit circle. Towards reformulating (29), consider the $p\left(\sum_{b \in \mathcal{B}} |\mathcal{N}_b|\right) \times 1$ supervector

$$\boldsymbol{\eta}(t) := \left[\left\{\boldsymbol{\eta}_{j'}^{b_1}(t)\right\}_{j' \in \mathcal{N}_{b_1}}^T \ldots \left\{\boldsymbol{\eta}_{j'}^{b_{|\mathcal{B}|}}(t)\right\}_{j' \in \mathcal{N}_{b_{|\mathcal{B}|}}}^T\right]^T \tag{31}$$

comprising the receiver noise of the bridge sensors' transmissions to their neighbors; i.e., the first $|\mathcal{N}_{b_1}|$ vectors in $\boldsymbol{\eta}(t)$ correspond to the reception noise at the neighbors of bridge sensor b_1 and so on. Using $\boldsymbol{\eta}(t)$, the noise supervectors $\boldsymbol{\eta}^{\alpha}(t)$ and $\boldsymbol{\eta}^{\beta}(t)$ can be written as $\boldsymbol{\eta}^{\alpha}(t) = \mathbf{P}_{\alpha}\boldsymbol{\eta}(t)$ and $\boldsymbol{\eta}^{\beta}(t) = \mathbf{P}_{\beta}\boldsymbol{\eta}(t)$, where the structure of the time-invariant matrices $\mathbf{P}_{\alpha}$ and $\mathbf{P}_{\beta}$ can be found in Appendix D, which establishes Lemma 3.

Lemma 3: *The LTV system in (29) can be equivalently written as*

$$\mathbf{y}(t+1) = \text{bdiag}\left(\mathbf{I}_{Jp}, \mu^{-\gamma/2}c_2\mathbf{A}\right)\mathbf{z}(t+1) + \begin{bmatrix} \mu^{1-\gamma}c_1\mathbf{I}_{Jp} \\ -\mu^{-\gamma/2}c_2\mathbf{I}_{Jp} \end{bmatrix}(\bar{\boldsymbol{\eta}}(t) + \boldsymbol{\eta}^{\alpha}(t)) \quad (32)$$

where the state $\mathbf{z}(t) := \left[\mathbf{z}_1^T(t)\ \ \mathbf{z}_2^T(t)\right]^T$ *is arbitrarily initialized as* $\mathbf{z}(0) := \left[\mathbf{y}_1^T(0)\ \ \left(\mathbf{y}_2'(0)\right)^T\right]^T$*; updated according to*

$$\begin{aligned}\mathbf{z}(t+1) = &\boldsymbol{\Phi}(t+1,\mu)\mathbf{z}(t) + \mathbf{R}_h^{\alpha}(t+1)\bar{\boldsymbol{\eta}}(t-1) \\ &+ \mathbf{R}_h^{\beta}(t+1)\boldsymbol{\eta}(t-1) \\ &+ [\boldsymbol{\epsilon}^T(t+1)\quad \mathbf{0}^T]^T \end{aligned} \quad (33)$$

and the transition matrix $\boldsymbol{\Phi}(t+1,\mu)$ *consists of the submatrices* $[\boldsymbol{\Phi}(t,\mu)]_{11} = \mathbf{I}_{Jp} - 2\mu\mathbf{R}_h(t) - \mu^{1-\gamma}c_1\mathbf{A}$, $[\boldsymbol{\Phi}(t,\mu)]_{12} = -\left(\mu^{1-\gamma}c_1/2\right)\mathbf{A}$ *and* $[\boldsymbol{\Phi}(t,\mu)]_{21} = [\boldsymbol{\Phi}(t,\mu)]_{22} = \mathbf{A}\mathbf{A}^{\dagger}$. *Matrices* $\mathbf{R}_h^{\alpha}(t+1)$ *and* $\mathbf{R}_h^{\beta}(t+1)$ *are defined as*

$$\mathbf{R}_h^{\alpha}(t+1) := \left[\frac{\mu^{1-\gamma}c_1}{2}\mathbf{I}_{Jp} - \mu^{2(1-\gamma)}c_1^2\mathbf{A}^T - 2\mu^{2-\gamma}c_1\mathbf{R}_h^T(t+1), \mu^{1-\gamma}c_1\mathbf{I}_{Jp}\right]^T \quad (34)$$

$$\mathbf{R}_h^{\beta}(t+1) := \left[\mu^{1-\gamma}c_1\left(\frac{3}{2}\mathbf{P}_{\alpha}^T - \mu^{1-\gamma}c_1(\mathbf{A}\mathbf{P}_{\alpha})^T - 2\mu(\mathbf{R}_h(t+1)\mathbf{P}_{\alpha})^T - \mathbf{P}_{\beta}^T\right), \mu^{1-\gamma}c_1\mathbf{P}_{\alpha}^T + 2\mathbf{C}_R^T\right]^T \quad (35)$$

with $\mathbf{C}_R$ *chosen such that* $\mathbf{A}\mathbf{C}_R = \mathbf{P}_{\beta} - \mathbf{P}_{\alpha}$.

A. Stability Analysis

This subsection deals with stability analysis of D-LMS based on the equivalent LTV system derived in Lemma 3. Specifically, it will be shown that under mild conditions the error norm $\|\mathbf{y}(t)\|$ remains most of the time in a finite interval, i.e., errors are weakly stochastic bounded (WSB) [20]. This WSB stability guarantees that for any $\theta > 0$ there exists a $\delta > 0$ such that $\Pr[\|\mathbf{y}(t)\| < \delta] = 1 - \theta$ uniformly in t. As a consequence, it will be shown that in the absence of observation and intersensor communication noise the local estimation errors in D-LMS converge exponentially fast to zero with probability one. Therefore, consensus is achieved a.s., as all local sensor estimates agree on the true parameter $\mathbf{s}_0$. This establishes a strong connection with the known behavior of C-LMS [20], further validating the importance of D-LMS in a distributed setting.

The first step in proving that $\|\mathbf{y}(t)\|$ is bounded in probability is to show that the same holds for $\|\mathbf{z}(t)\|$. This will be established under the following assumptions:

a3) *The regressor vectors* $\mathbf{h}_j(t)$ *are strictly stationary with* $\mathbf{R}_{h_j} := E\left[\mathbf{h}_j(t)\mathbf{h}_j^T(t)\right] < \infty$, *and* $\mathbf{R}_h := \text{bdiag}(\mathbf{R}_{h_1}, \ldots, \mathbf{R}_{h_J}) > \mathbf{0}$. *Regressors are ergodic and satisfy (a.s.)*

$$\lim_{t\to\infty} t^{-1}\sum_{\tau=1}^{t}\|\mathbf{R}_h - \mathbf{R}_h(\tau)\| = E[\|\mathbf{R}_h - \mathbf{R}_h(t)\|] := m_{\tilde{h}} < \infty. \quad (36)$$

a4) *Communication and observation noise vector norms are bounded in the mean, i.e.,*

$$E[\|\epsilon_j(t)\|] = m_{\epsilon} < \infty, \quad E[\|\boldsymbol{\eta}_j^b(t)\|] = m_{\eta} < \infty, \quad \text{and } E[\|\bar{\boldsymbol{\eta}}_b^j(t)\|] = m_{\bar{\eta}} < \infty. \quad (37)$$

Under assumption a4), typically met in practice, the estimation error does not grow unbounded.

Necessary for proving boundedness of $\mathbf{z}(t)$ in (33) is to show that the norm of $\boldsymbol{\Phi}(1:t,\mu) := \prod_{\tau=1}^{t}\boldsymbol{\Phi}(\tau,\mu)$ converges to zero as $t \to \infty$, a.s.. To this end, let $\lambda^{+}(\mathbf{A})$ denote the minimum positive eigenvalue of $\mathbf{A} \geq \mathbf{0}$. Specifically, it is established in Appendix E that:

Lemma 4: *If a3) holds true,* c_1 *is selected such that* $c_1\lambda^{+}(\mathbf{A}) > 4$ *and the step-size* $\mu \in (0, \mu_u')$*, with* $\mu_u' < \mu_u$ *where*

$$\mu_u := \left[2\min\left(\lambda_{\max}^{-1}\left(\mathbf{R}_h + \frac{c_1}{2}\mathbf{A}\right), \lambda_{\max}^{-1}\left(2\mathbf{R}_h + \frac{3c_1}{4}\mathbf{A}\right)\right)\right]^{\frac{1}{1-\gamma}} \quad (38)$$

is chosen to guarantee that the eigenvalues of $E_{\mathbf{h}}[\boldsymbol{\Phi}(t,\mu)]$ *are less than one, then with* $\lambda \in (0,1)$

$$\limsup_{t\to\infty} t^{-1}\log\|\boldsymbol{\Phi}(1:t,\mu)\| \leq \log(\lambda) < 0, \text{ a.s.} \quad (39)$$

In words, (39) establishes that $\|\boldsymbol{\Phi}(1:t,\mu)\|$ will converge to zero exponentially fast with probability one. This property is necessary to prove later on that $\|\mathbf{z}(t)\|$ satisfies the WSB property, which in turn will lead us to the ultimate goal of establishing stochastic boundedness of $\|\mathbf{y}(t)\|$. Thus, using Lemma 4, we prove in Appendix F the following result.

Lemma 5: *Under a3)–a4),* $c_1\lambda^{+}(\mathbf{A}) > 4$ *and if* $\mu \in (0, \mu_u')$ *with* $\mu_u' < \mu_u$*, then* $\mathbf{z}(t)$ *satisfies the WSB property, i.e.,*

$$\lim_{\delta\to\infty}\sup_{t\geq 0} Pr[\|\mathbf{z}(t)\| \geq \delta] = 0. \quad (40)$$

Careful inspection of (32), and exploitation of Lemmas 4 and 5, along with a4), reveals that $\mathbf{y}(t)$ is also WSB. As a result, it is possible to prove the following main result (see Appendix G).

Proposition 1: *Under a2)–a4),* $c_1\lambda^{+}(\mathbf{A}) > 4$ *and if* $\mu \in (0, \mu_u')$ *with* $\mu_u' < \mu_u$*, then* $\mathbf{y}(t)$ *is WSB; i.e.,*

$$\lim_{\delta\to\infty}\sup_{t\geq 0} Pr[\|\mathbf{y}(t)\| \geq \delta] = 0. \quad (41)$$

Proposition 1 asserts there is no probability mass of $\mathbf{y}(t)$, allowing local estimation errors $\mathbf{y}_{1,j}(t) := \mathbf{s}_j(t) - \mathbf{s}_0$ escape to infinity. This is very important since even in the presence of observation and communication noise the local estimation errors remain bounded. Interestingly, in the absence of noise the local estimates $\mathbf{s}_j(t)$ provided by D-LMS converge exponentially fast

to $\mathbf{s}_0$ with probability one. Thus, D-LMS exhibits behavior similar to its centralized counterpart when it comes to stationary ergodic signals [20]. Actually, it follows readily from Lemma 5 that (cf. [19, Proof of Lemma 3]):

Corollary 1: *If $c_1\lambda^+(\mathbf{A}) > 4$ and $\mu \in (0, \mu_u')$ with $\mu_u' < \mu_u$, a2) and a3) hold true, while $\epsilon_j(t) = 0$ and $\boldsymbol{\eta}_j^b(t) = \bar{\boldsymbol{\eta}}_b^j(t) = \mathbf{0}$ for $j \in \mathcal{J}$ and $b \in \mathcal{B}$, then there exists $t_0 < \infty$, a random variable $B < \infty$ and $\lambda(\mu) \in (0,1)$ such that*

$$\|\mathbf{y}(t)\| \leq B\lambda^t(\mu), \text{ a.s.} \quad \forall\, t > t_0. \tag{42}$$

Corollary 1 demonstrates that the WSN achieves consensus in the sense that local estimation errors $\{\mathbf{s}_j(t) - \mathbf{s}_0\}_{j=1}^J$ converge to zero exponentially fast on a per realization basis. Interestingly, μ_u resembles the stability bound for C-LMS, namely $2/\lambda_{\max}(\mathbf{R}_h)$ [21, Ch. 9]. The main difference here is that this bound is also affected by the topology of the WSN, via $\mathbf{A}$, due to the distributed nature of the algorithm and the information exchanges among sensors.

B. Performance Analysis

In this subsection, the estimation performance of D-LMS is analyzed by approximating the error covariance matrix. Since the recursion (32) involved in D-LMS is time-varying, a closed-form expression for the error covariance matrix is difficult, if not impossible, to obtain. Specifically, the estimation MSE associated with the D-LMS recursions in (32)–(33) evolves according to

$$\begin{aligned} e_{\text{mse}}(t) :=& J^{-1}E\left[\mathbf{y}_1^T(t)\mathbf{y}_1(t)\right] = J^{-1}E\left[\mathbf{z}_1^T(t)\mathbf{z}_1(t)\right] \\ &+ \left(\mu^{1-\gamma}c_1\right)^2 J^{-1}\text{trace}\Big(\mathbf{R}_{\bar{\eta}} + \mathbf{P}_\alpha\mathbf{R}_\eta\mathbf{P}_\alpha^T\Big) \end{aligned} \tag{43}$$

where $E\left[\mathbf{z}_1^T(t)\mathbf{z}_1(t)\right]$ denotes the trace of the upper left $Jp\times Jp$ submatrix of the covariance matrix $E[\mathbf{z}(t)\mathbf{z}^T(t)]$ that evolves according to

$$\begin{aligned} E[\mathbf{z}(t)\mathbf{z}^T(t)] =& E\Big[\boldsymbol{\Phi}(t,\mu)\mathbf{z}(t-1)\mathbf{z}^T(t-1)\boldsymbol{\Phi}^T(t,\mu)\Big] \\ &+ E\Big[\mathbf{R}_h^\alpha(t)\mathbf{R}_{\bar{\eta}}\left(\mathbf{R}_h^\alpha(t)\right)^T\Big] + E\Big[\mathbf{R}_h^\beta(t)\mathbf{R}_\eta\left(\mathbf{R}_h^\beta(t)\right)^T\Big] \\ &+ \text{bdiag}\left(E\left[\boldsymbol{\epsilon}(t)\boldsymbol{\epsilon}^T(t)\right], \mathbf{0}_{Jp\times Jp}\right) \\ &+ E\left[\boldsymbol{\Phi}(t,\mu)\mathbf{z}(t-1)\left[\boldsymbol{\epsilon}^T(t)\ \ \mathbf{0}^T\right]^T\right] \\ &+ \left(E\left[\boldsymbol{\Phi}(t,\mu)\mathbf{z}(t-1)[\boldsymbol{\epsilon}^T(t)\ \ \mathbf{0}^T]^T\right]\right)^T. \end{aligned} \tag{44}$$

The first expectation in the rhs of (44) is impossible to split because the regressors are temporally correlated. Thus, under a3)–a4) it appears impossible to evaluate $E[\mathbf{z}(t)\mathbf{z}^T(t)]$. One possible alternative is to consider an appropriate time-invariant "average" system approximating the LTV system in (32) and recursively evaluate its corresponding error covariance matrix. Then, using stochastic averaging arguments, see, e.g., [21, Ch. 9], the estimation error $\bar{\mathbf{y}}_1(t)$ associated with the "average" system can be shown convergent in probability to $\mathbf{y}_1(t)$ as the step-size μ approaches zero. This approach allows approximating the estimation MSE of D-LMS with that of the average system

$$\begin{aligned} \bar{\mathbf{y}}(t+1) =& \text{bdiag}\left(\mathbf{I}_{Jp}, \mu^{-\gamma/2}c_2\mathbf{A}\right)\bar{\mathbf{z}}(t+1) \\ &+ \begin{bmatrix} \mu^{1-\gamma}c_1\mathbf{I}_{Jp} \\ -\mu^{-\gamma/2}c_2\mathbf{I}_{Jp} \end{bmatrix}\left(\bar{\boldsymbol{\eta}}(t) + \boldsymbol{\eta}^\alpha(t)\right) \end{aligned} \tag{45}$$

$$\begin{aligned} \bar{\mathbf{z}}(t+1) =& \boldsymbol{\Phi}(\mu)\bar{\mathbf{z}}(t) + \mathbf{R}_h^\alpha\bar{\boldsymbol{\eta}}(t-1) \\ &+ \mathbf{R}_h^\beta\boldsymbol{\eta}(t-1) + [\,\boldsymbol{\epsilon}^T(t+1)\quad \mathbf{0}^T\,]^T \end{aligned} \tag{46}$$

where $\boldsymbol{\Phi}(\mu) := E_{\mathbf{h}}[\boldsymbol{\Phi}(t,\mu)]$, $\mathbf{R}_h^\alpha := E_{\mathbf{h}}[\mathbf{R}_h^\alpha(t+1)]$ and $\mathbf{R}_h^\beta := E_{\mathbf{h}}[\mathbf{R}_h^\beta(t+1)]$. Note that the average system in (45) is not constructed by taking expectations on both sides of (32)–(33). Instead, it is formed starting from the primary system and replacing the time-varying transition matrix $\boldsymbol{\Phi}(t,\mu)$, and the matrices $\mathbf{R}_h^\alpha(t+1)$, $\mathbf{R}_h^\beta(t+1)$ with their time-invariant counterparts $\boldsymbol{\Phi}(\mu)$, and $\mathbf{R}_h^\alpha$, $\mathbf{R}_h^\beta$ respectively. This average system plays a key role in the stochastic averaging approach of [21, Ch. 9].

Next, we see how the local estimation errors in $\mathbf{y}_1(t)$ are statistically related with the average state vector $\bar{\mathbf{y}}_1(t)$. Recall that both $\mathbf{y}_1(t)$ and $\bar{\mathbf{y}}_1(t)$ depend on μ. Actually, it is shown in Appendix H that:

Proposition 2: *If $\mathbf{z}(0) = \bar{\mathbf{z}}(0)$ and a2)–a4) are satisfied, while the joint moments of $\{\mathbf{h}_j(t)\}_{j=1}^J$ are bounded, then given finite $T > 0$ and for any $\delta > 0$ and $\beta > 0$ arbitrarily small, it holds that*

$$\lim_{\mu\to 0}\sup_{0\leq t\leq T/\mu^{1-\gamma-\beta}} Pr[\|\bar{\mathbf{y}}_1(t) - \mathbf{y}_1(t)\| \geq \delta] = 0. \tag{47}$$

Proposition 2 shows that the probability of the estimation error $\mathbf{y}_1(t)$ being close to $\bar{\mathbf{y}}_1(t)$ approaches unity with vanishing step-sizes, so long as the D-LMS and its "average" version are initialized with the same local estimates and multipliers, i.e., $\mathbf{z}(0) = \bar{\mathbf{z}}(0)$. This type of result is referred to as *trajectory locking*, because the trajectory of the primary system hovers around and locks to the trajectory of its average counterpart. The time horizon for which the two systems remain "locked" goes to infinity as $\mu \to 0$. These locking results are applicable even when regressors exhibit temporal correlations.

Now, observe that the "average" D-LMS algorithm in (45) has a time-invariant transition matrix. As a result, the "average" estimation error covariance $\mathbf{R}_{\bar{y}_1}(t) := E[\bar{\mathbf{y}}_1(t)\bar{\mathbf{y}}_1^T(t)]$ can be found in closed form. Specifically, we prove in Appendix I that

$$\begin{aligned} \mathbf{R}_{\bar{y}_1}(t+1) =& [\mathbf{R}_{\bar{z}}(t+1)]_{11} \\ &+ \left(\mu^{1-\gamma}c_1\right)^2\left(\mathbf{R}_{\bar{\eta}} + \mathbf{P}_\alpha\mathbf{R}_\eta\mathbf{P}_\alpha^T\right) \end{aligned} \tag{48}$$

where $[\mathbf{R}_{\bar{z}}(t+1)]_{11}$ is the $Jp \times Jp$ upper left submatrix of

$$\begin{aligned} \mathbf{R}_{\bar{z}}(t+1) =& \boldsymbol{\Phi}^{t+1}(\mu)\mathbf{R}_{\bar{z}}(0)(\boldsymbol{\Phi}^{t+1}(\mu))^T \\ &+ \sum_{\tau=0}^{t}\boldsymbol{\Phi}^\tau(\mu)\mathbf{R}_{\check{\eta}}(\boldsymbol{\Phi}^\tau(\mu))^T \end{aligned} \tag{49}$$

while

$$\begin{aligned} \mathbf{R}_{\check{\eta}} =& \mathbf{R}_h^\alpha\mathbf{R}_{\bar{\eta}}\left(\mathbf{R}_h^\alpha\right)^T + \mathbf{R}_h^\beta\mathbf{R}_\eta\left(\mathbf{R}_h^\beta\right)^T \\ &+ \text{bdiag}\left(E[\boldsymbol{\epsilon}(t+1)\boldsymbol{\epsilon}^T(t+1)], \mathbf{0}_{Jp\times Jp}\right) \end{aligned} \tag{50}$$

with $\mathbf{R}_{\bar{\eta}} := E[\bar{\boldsymbol{\eta}}(t)\bar{\boldsymbol{\eta}}^T(t)]$ and $\mathbf{R}_\eta := E[\boldsymbol{\eta}(t)\boldsymbol{\eta}^T(t)]$; see also Appendix I for the structure of $\mathbf{R}_{\bar{\eta}}$ and $\mathbf{R}_\eta$.

Based on Proposition 2, the covariance matrix in (48) can be viewed as an approximation to $\mathbf{R}_{y_1}(t) := E\left[\mathbf{y}_1(t)\mathbf{y}_1^T(t)\right]$. Then, the global normalized estimation error of D-LMS at time instant t can be approximated as

$$e_{\text{mse}}(t) := \frac{1}{J}\sum_{j=1}^{J}E\left[\|\mathbf{s}_j(t) - \mathbf{s}_0\|^2\right] \approx J^{-1}\text{trace}(\mathbf{R}_{\bar{y}_1}(t)) \tag{51}$$

while local approximate MSE performance across sensors can be acquired from the corresponding diagonal entries of $\mathbf{R}_{\bar{y}_1}(t)$. Proposition 2 implies that this approximation improves as $\mu \to 0$. Intuitively, this happens because a vanishing step-size suppresses temporal correlations present in the regressors thus making D-LMS behave as the "average system" in (45). In the interest of tractability, the "average system" does not take into account temporal correlations. Thus, for a small step-size the MSE corresponding to the "average" system in (45) can be used to approximate efficiently the one associated with D-LMS, which will be corroborated via simulations.

Remark 7 (Comparison With Existing Results): The stochastic stability results presented in Section IV-A allow for (non-) Gaussian distributed and spatio-temporally correlated regressors. As for C-LMS, WSB has been established under conditions similar to those in [20]. Communication noise has not been considered earlier, and stochastic boundedness of D-LMS is a consequence of the inherent noise robustness of the method of multipliers; see also [18] for related claims in single-shot nonadaptive distributed estimation. Mean and MSE convergence results for diffusion LMS [9] were established under the widely assumed white Gaussian setting [21, Ch. 5]; similar mean-stability results were reported for D-LMS in [10]. A comparison between the stability results here and those in [9] is not possible since they are different in nature. Regarding performance analysis, steady-state closed-form expressions of the relevant figures of merit have been derived for both incremental and diffusion LMS [8], [9]; when regressors are Gaussian and independent in space and time by relying on an energy conservation framework [16]. This should be contrasted with the alternative approach delineated in Section IV-B, that utilizes stochastic averaging arguments to approximate the MSE associated with D-LMS. This approximation becomes increasingly accurate for a vanishing step-size, since the regressor's temporal correlations are suppressed making D-LMS behave as in a white data setting.

C. Numerical Example

Here again, we test a WSN generated as a $\mathcal{G}^2(80, 0.6)$ graph yielding $\lambda^+(\mathbf{A}) = 0.904$. With $p = 8$ and $\mathbf{s}_0 = \mathbf{1}_p$, observations obey the linear model (21), where regressors are $\mathbf{h}_j(t) = [h_j(t) \dots h_j(t-p+1)]^T$ with $h_j(t)$ evolving according to an AR(1) process as in Section III-C. We choose $\rho = 0.7$, the $u_{1,j} \sim \mathcal{U}[0,1]$ are i.i.d. in space and the uniformly distributed white driving noise has a spatial variance profile given by $\sigma^2_{\nu_j} = 2u_{2,j}$ with $u_{2,j} \sim \mathcal{U}[0,1]$ and i.i.d.. First, we corroborate the result in Corollary 1, by running D-LMS with $c_j = 2.25$, $\varepsilon_j = \mu^{-1/4}$ for all $j \in \mathcal{J}$ in a noise-free setup and computing the sample paths of the normalized estimation error $J^{-1}\sum_{j=1}^{J} \|\mathbf{s}_j(t) - \mathbf{s}_0\|$. Results are depicted in Fig. 6 (top) for different step-sizes related to the upper bound $\mu_u = 8 \times 10^{-3}$. When $\mu \in (0, \mu_u)$, the error norm converges to zero exponentially fast with a decay rate increasing with μ. As per Corollary 1, convergence cannot be claimed for step-sizes larger than μ_u, though simulations indicate that the stability region may be larger than $(0, \mu_u)$. Next, we validate the approximation (51) by plotting the empirically estimated MSE achieved by D-LMS in (32) (averaged over 50 Monte Carlo runs) and comparing it with the MSE achieved by the average system [cf. rhs of (51)] in a $\mathcal{G}^2(60, 0.25)$ setting. Space-time i.i.d. observation noise $\epsilon_j(t) \sim \mathcal{N}(0, 10^{-3})$ is now added as well as receiver AWGN of variance $\sigma^2_\eta = 10^{-2}$. Fig. 6 (bottom) confirms that the theoretical MSE obtained from the "average" D-LMS in (45) approximates well the primary MSE in (32), especially as μ becomes smaller. Note that D-LMS is derived using the AD-MoM. Convergence in MoM is not necessarily monotonic and the same holds for the true and approximated "average" behavior of D-LMS, as depicted in Fig. 6 (bottom).

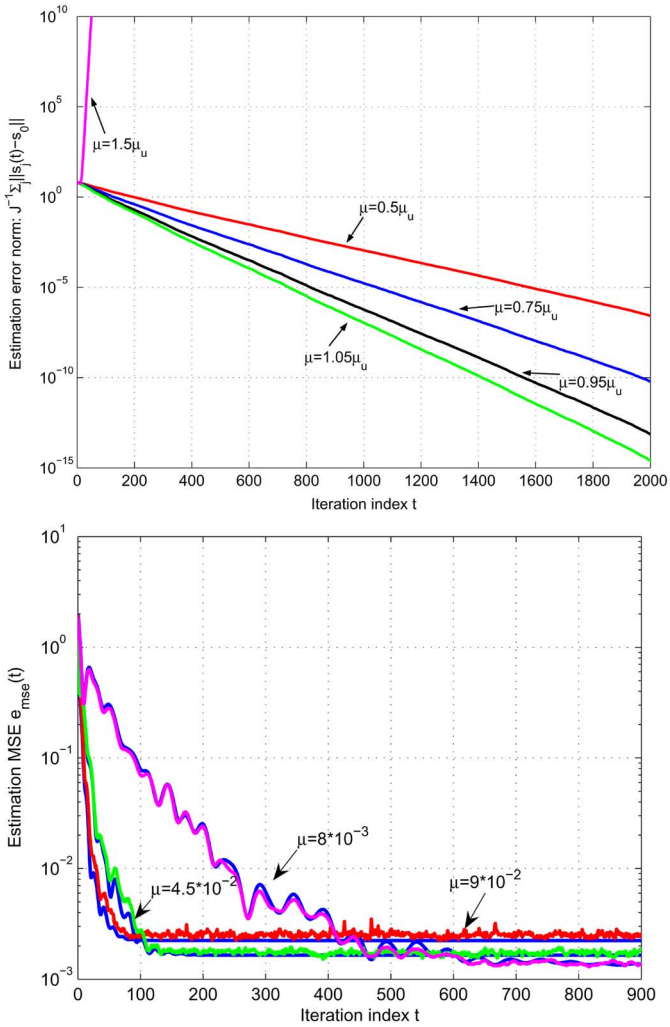

Fig. 6. (top) Normalized estimation error for D-LMS in the absence of noise; (bottom) Empirical normalized estimation MSE for D-LMS and theoretical approximation (51) for the "average" D-LMS.

V. Concluding Remarks

We developed a distributed LMS-type of adaptive algorithm for WSN based tracking applications, where intersensor communications are constrained to single-hop neighboring sensors and are challenged by the effects of additive receiver noise. Starting from a well-posed convex optimization problem defining the desired estimator, we reformulated it into an equivalent constrained form whose structure lends itself naturally to decentralized implementation. We capitalized on this favorable structure by resorting to the alternating-direction method of multipliers, while using stochastic approximation tools in the process we finally arrived at simple recursions. The resulting in-network processing per sensor was interpreted

as a local-LMS adaptation rule superimposed to the output of a tunable PI regulator, which drives the local estimate to consensus as dictated by a network-wide information enriched reference. Numerical examples illustrated that D-LMS outperforms comparable adaptive schemes, and has the potential of tracking nonstationary processes.

The challenging problem of algorithm stability in a stochastic sense has been also addressed. The distributed stability results obtained are closely related to those available for the centralized setting. For observations adhering to a linear model, stationary ergodic regressors, and a fixed step-size below a positive threshold, we established that D-LMS incurs local estimation errors satisfying the WSB property even in the presence of additive intersensor communication noise. In the absence of noise, D-LMS estimates were shown a.s. exponentially convergent to the true parameter of interest. With regards to performance analysis, we established a stochastic trajectory locking result which shows that for small step-sizes, the D-LMS estimation error trajectories closely follow the ones of its time-invariant "averaged" system mate. An "averaged" system estimation error covariance matrix was obtained in closed form, that provided a means of accurately approximating the actual D-LMS estimation MSE as corroborated by numerical simulations.[2]

APPENDIX

A. Proof of Lemma 1

Recall that in the presence of noise, bridge variables obey [cf. (16)]

$$\bar{\mathbf{s}}_b(t+1)=\sum_{j\in\mathcal{N}_b}\frac{\mathbf{v}_j^b(t)+\mu^{-\frac{\gamma}{2}}c_2\left(\mathbf{s}_j(t+1)+\bar{\boldsymbol{\eta}}_b^j(t+1)\right)}{|\mathcal{N}_b|\mu^{-\frac{\gamma}{2}}c_2}. \quad (52)$$

Substituting (14) into (52), while adding and subtracting $\sum_{j\in\mathcal{N}_b}\bar{\boldsymbol{\eta}}_b^j(t)$, yields

$$\begin{aligned}\bar{\mathbf{s}}_b(t+1)&=\sum_{j\in\mathcal{N}_b}\frac{\mathbf{v}_j^b(t-1)+\mu^{-\frac{\gamma}{2}}c_2\left(\mathbf{s}_j(t)+\bar{\boldsymbol{\eta}}_b^j(t)\right)}{|\mathcal{N}_b|\mu^{-\frac{\gamma}{2}}c_2}\\&\quad-\bar{\mathbf{s}}_b(t)+|\mathcal{N}_b|^{-1}\sum_{j\in\mathcal{N}_b}\mathbf{s}_j(t+1)\\&\quad+|\mathcal{N}_b|^{-1}\sum_{j\in\mathcal{N}_b}\left[\bar{\boldsymbol{\eta}}_b^j(t+1)-\bar{\boldsymbol{\eta}}_b^j(t)\right]-|\mathcal{N}_b|^{-1}\sum_{j\in\mathcal{N}_b}\boldsymbol{\eta}_j^b(t). \quad (53)\end{aligned}$$

Equations (22) and (52) imply that the first sum in (53) equals $\bar{\mathbf{s}}_b(t)$ for $t\geq 0$. Thus, (23) follows. □

B. Proof of Lemma 2

Summing (14) over $b\in\mathcal{B}_j$, we can write $\mathbf{y}_{2,j}(t+1)$ as

$$\begin{aligned}\mathbf{y}_{2,j}(t+1)&=\sum_{b\in\mathcal{B}_j}\Big[\mathbf{v}_j^b(t-1)+\mu^{-\gamma/2}c_2\big(\mathbf{s}_j(t)\\&\qquad-\left(\bar{\mathbf{s}}_b(t)+\boldsymbol{\eta}_j^b(t)\right)\big)\Big]\\&=\mathbf{y}_{2,j}(t)+\mu^{-\gamma/2}c_2|\mathcal{B}_j|\mathbf{s}_j(t)\\&\quad-\mu^{-\gamma/2}c_2\sum_{b\in\mathcal{B}_j}\left(\sum_{j'\in\mathcal{N}_b}\frac{\mathbf{s}_{j'}(t)}{|\mathcal{N}_b|}\right)\\&\quad-\mu^{-\gamma/2}c_2(\bar{\boldsymbol{\eta}}_j(t)-\bar{\boldsymbol{\eta}}_j(t-1))\\&\quad-\mu^{-\gamma/2}c_2\left(\boldsymbol{\eta}_j^{\alpha}(t)-\boldsymbol{\eta}_j^{\beta}(t-1)\right) \quad (54)\end{aligned}$$

where the last equality follows after splitting the sum in the first equality into four individual terms and invoking Lemma 1. Adding and subtracting $\mu^{-\gamma/2}c_2|\mathcal{B}_j|\mathbf{s}_0$ from the rhs of (54), yields (27) readily.

To prove (26), recall that the local estimate $\mathbf{s}_j(t)$ is updated as [cf. (15)]

$$\begin{aligned}\mathbf{s}_j(t+1)&=\mathbf{s}_j(t)+2\mu\mathbf{h}_j(t+1)e_j(t+1)\\&\quad-\frac{\mu^{1-\gamma}c_1}{2}\Bigg[|\mathcal{B}_j|\mathbf{s}_j(t)+\frac{\mu^{\gamma/2}}{c_2}\mathbf{y}_{2,j}(t+1)\\&\qquad-\sum_{b\in\mathcal{B}_j}\bar{\mathbf{s}}_b(t)-\sum_{b\in\mathcal{B}_j}\boldsymbol{\eta}_j^b(t)\Bigg]. \quad (55)\end{aligned}$$

Upon i) using $e_j(t):=x_j(t)-\mathbf{h}_j^T(t)\mathbf{s}_j(t-1)$, ii) substituting $x_j(t)$ from a2) into (55), iii) subtracting $\mathbf{s}_0$ from both sides of (55), and iv) replacing $\bar{\mathbf{s}}_b(t)$ and $\mathbf{y}_{2,j}(t+1)$ from (23) and (27), respectively, we arrive at

$$\begin{aligned}&\mathbf{y}_{1,j}(t+1)\\&\quad=\mathbf{y}_{1,j}(t)-2\mu\mathbf{h}_j(t+1)\mathbf{h}_j^T(t+1)\mathbf{y}_{1,j}(t)-\mu^{1-\gamma}c_1\\&\qquad\times\Bigg[|\mathcal{B}_j|\mathbf{s}_j(t)+\frac{\mu^{\gamma/2}}{2c_2}\mathbf{y}_{2,j}(t)-\sum_{b\in\mathcal{B}_j}\left(\sum_{j'\in\mathcal{N}_b}\frac{\mathbf{s}_{j'}(t)}{|\mathcal{N}_b|}\right)\\&\qquad\quad-(\bar{\boldsymbol{\eta}}_j(t)-\bar{\boldsymbol{\eta}}_j(t-1))-\left(\boldsymbol{\eta}_j^{\alpha}(t)-\boldsymbol{\eta}_j^{\beta}(t-1)\right)\Bigg]\\&\qquad+2\mu\mathbf{h}_j(t+1)\epsilon_j(t+1)\\&\quad=\mathbf{y}_{1,j}(t)-2\mu\mathbf{h}_j(t+1)\mathbf{h}_j^T(t+1)\mathbf{y}_{1,j}(t)-\mu^{1-\gamma}c_1\\&\qquad\times\Bigg[|\mathcal{B}_j|\mathbf{y}_{1,j}(t)-\sum_{b\in\mathcal{B}_j}\left(\sum_{j'\in\mathcal{N}_b}\frac{\mathbf{y}_{1,j'}(t)}{|\mathcal{N}_b|}\right)\\&\qquad\quad+\frac{\mu^{\gamma/2}}{2c_2}\mathbf{y}_{2,j}(t)-(\bar{\boldsymbol{\eta}}_j(t)-\bar{\boldsymbol{\eta}}_j(t-1))\\&\qquad\quad-\left(\boldsymbol{\eta}_j^{\alpha}(t)-\boldsymbol{\eta}_j^{\beta}(t-1)\right)\Bigg]+2\mu\mathbf{h}_j(t+1)\epsilon_j(t+1)\end{aligned}$$

where the last equality follows after adding and subtracting $|\mathcal{B}_j|\mathbf{s}_0$ from the quantity inside the square brackets in the rhs of the first equality. □

C. Proof of (29)

Consider first the noise vectors in (29). Stacking the channel noise terms from (26) and (27) and scaling with $\mu^{1-\gamma}c_1$ and $\mu^{-\gamma/2}c_2$, respectively, yields the first noise term in (29). Likewise, stacking the noise terms $2\mu\mathbf{h}_j(t+1)\epsilon_j(t+1)$ in (26) for $j=1,\ldots,J$ yields the second noise term in (29) corresponding to the observation noise. Note that (27) contains no observation noise, which explains the zero vector at the lower part of the second noise term in (29).

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

The second term within the square brackets in (26) explains why $[\mathbf{\Psi}(t,\mu)]_{12} = -\left(\mu^{1-\gamma/2}c_1/(2c_2)\right)\mathbf{I}_{Jp}$. To specify the structure of $[\mathbf{\Psi}(t,\mu)]_{11}$, notice that

$$\sum_{b\in\mathcal{B}_j}|\mathcal{N}_b|^{-1}\sum_{j'\in\mathcal{N}_b}\mathbf{y}_{1,j'}(t)=\sum_{b\in\mathcal{B}}\mathbf{e}_b(j)|\mathcal{N}_b|^{-1}(\mathbf{e}_b\otimes\mathbf{I}_p)^T\mathbf{y}_1(t). \tag{56}$$

The supervector formed by concatenating the first term within the square brackets in (26), for $j=1,\ldots,J$ can be written as

$$\sum_{b\in\mathcal{B}}\frac{(\mathbf{e}_b\otimes\mathbf{I}_p)(\mathbf{e}_b\otimes\mathbf{I}_p)^T}{|\mathcal{N}_b|}\mathbf{y}_1(t). \tag{57}$$

Stack the first term in (26) for $j=1,\ldots,J$ and add the resulting supervector to the one in (57), to obtain

$$\begin{aligned}&\Bigg(\mathbf{I}_{Jp}-2\mu\mathbf{R}_h(t+1)-\mu^{1-\gamma}c_1\Bigg[\Big(\mathrm{bdiag}(|\mathcal{B}_1|\mathbf{I}_p,\ldots,|\mathcal{B}_J|\mathbf{I}_p)\\&\quad-\sum_{b\in\mathcal{B}}|\mathcal{N}_b|^{-1}(\mathbf{e}_b\otimes\mathbf{I}_p)(\mathbf{e}_b\otimes\mathbf{I}_p)^T\Big)\Bigg]\Bigg)\mathbf{y}_1(t)\\&\quad=\left(\mathbf{I}_{Jp}-2\mu\mathbf{R}_h(t+1)-\mu^{1-\gamma}c_1\mathbf{A}\right)\mathbf{y}_1(t)\end{aligned}$$

from which we can readily deduce that $[\mathbf{\Psi}(t,\mu)]_{11}$ is equal to the matrix multiplying $\mathbf{y}_1(t)$. Also, it follows immediately from the first term in (27) that $[\mathbf{\Psi}(t,\mu)]_{22}=\mathbf{I}_{Jp}$. Further, note that the second term within the square brackets in (27) has the same structure as the second term in (26). Thus, after i) stacking the first and second terms within the square brackets in (27), scaling them with $\mu^{-\gamma/2}c_2$ and subtracting them and ii) using (56) and (57), we obtain $[\mathbf{\Psi}(t,\mu)]_{21}=\mu^{-\gamma/2}c_2\mathbf{A}$. □

D. Proof of Lemma 3

We will argue by induction. For $t=0$ we have from (33) that $\mathbf{z}(1)=\mathbf{\Phi}(1,\mu)\mathbf{z}(0)+\left[\boldsymbol{\epsilon}^T(1)\ \mathbf{0}^T\right]^T$, where $\mathbf{z}(0):=\left[\mathbf{y}_1^T(0)\ \ (\mathbf{y}_2'(0))^T\right]^T$; and after substituting $\mathbf{z}(1)$ into (32), we find

$$\begin{aligned}\mathbf{y}(1)&=\mathrm{bdiag}\left(\mathbf{I}_{Jp},\mu^{-\gamma/2}c_2\mathbf{A}\right)\mathbf{\Phi}(1,\mu)\mathbf{z}(0)\\&\quad+\begin{bmatrix}\mu^{1-\gamma}c_1\mathbf{I}_{Jp}\\-\mu^{-\gamma/2}c_2\mathbf{I}_{Jp}\end{bmatrix}(\bar{\boldsymbol{\eta}}(0)+\boldsymbol{\eta}^{\alpha}(0))+\begin{bmatrix}\boldsymbol{\epsilon}(1)\\\mathbf{0}\end{bmatrix}.\end{aligned} \tag{58}$$

Note that

$$\begin{aligned}&\mathrm{bdiag}\left(\mathbf{I}_{Jp},\mu^{-\gamma/2}c_2\mathbf{A}\right)\mathbf{\Phi}(t,\mu)\\&\qquad=\mathbf{\Psi}(t,\mu)\mathrm{bdiag}\left(\mathbf{I}_{Jp},\mu^{-\gamma/2}c_2\mathbf{A}\right)\end{aligned}$$

for $t\geq 1$; and $\mathbf{y}(0)=\mathrm{bdiag}\left(\mathbf{I}_{Jp},\mu^{-\gamma/2}c_2\mathbf{A}\right)\mathbf{z}(0)$. Thus, the rhs of (58) is equal to the rhs of (29) for $t=0$.

Suppose next that (32) and (33) hold true for $\mathbf{y}(t)$ and $\mathbf{z}(t)$. The same will be shown for $\mathbf{y}(t+1)$ and $\mathbf{z}(t+1)$. To this end, replace $\mathbf{y}(t)$ with the rhs of (32) evaluated at time instant t, into (29) to obtain

$$\begin{aligned}&\mathbf{y}(t+1)\\&\quad=\mathrm{bdiag}\left(\mathbf{I},\mu^{-\gamma/2}c_2\mathbf{A}\right)\mathbf{\Phi}(t+1,\mu)\mathbf{z}(t)\\&\qquad+(\mathbf{\Psi}(t+1,\mu)-\mathbf{I}_{Jp})\begin{bmatrix}\mu^{1-\gamma}c_1\mathbf{I}_{Jp}\\-\mu^{-\gamma/2}c_2\mathbf{I}_{Jp}\end{bmatrix}\bar{\boldsymbol{\eta}}(t-1)\\&\qquad+\mathbf{\Psi}(t+1,\mu)\begin{bmatrix}\mu^{1-\gamma}c_1\mathbf{I}_{Jp}\\-\mu^{-\gamma/2}c_2\mathbf{I}_{Jp}\end{bmatrix}\mathbf{P}_{\alpha}\boldsymbol{\eta}(t-1)\end{aligned} \tag{59}$$

$$-\begin{bmatrix}\mu^{1-\gamma}c_1\mathbf{I}_{Jp}\\-\mu^{-\gamma/2}c_2\mathbf{I}_{Jp}\end{bmatrix}\mathbf{P}_{\beta}\boldsymbol{\eta}(t-1) \tag{60}$$

$$+\begin{bmatrix}\mu^{1-\gamma}c_1\mathbf{I}_{Jp}\\-\mu^{-\gamma/2}c_2\mathbf{I}_{Jp}\end{bmatrix}(\bar{\boldsymbol{\eta}}(t)+\boldsymbol{\eta}^{\alpha}(t))+\begin{bmatrix}\boldsymbol{\epsilon}(t+1)\\\mathbf{0}\end{bmatrix} \tag{61}$$

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

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

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

and $\boldsymbol{\nu}_{|\mathcal{N}_{b_r}|,r(j)}$ denotes the $|\mathcal{N}_{b_r}|\times 1$ vector having all entries equal to zero except the $r(j)$th entry which is unity. Note that $r(j)=1,\ldots,|\mathcal{N}_{b_r}|$ denotes the order in which $\boldsymbol{\eta}_j^{b_r}(t)$ appears in $\left\{\boldsymbol{\eta}_{j'}^{b_r}(t)\right\}_{j'\in\mathcal{N}_{b_r}}$.

Coming back to prove that $\mathbf{y}(t+1)$ is updated according to Lemma 3, observe that the noise term in (59) can be easily written as $\mathrm{bdiag}\left(\mathbf{I}_{Jp},\mu^{-\gamma/2}c_2\mathbf{A}\right)\mathbf{R}_h^{\alpha}(t+1)\bar{\boldsymbol{\eta}}(t-1)$. Then, after algebraic manipulations the noise terms in (60) can be expressed as $\check{\mathbf{R}}_h^{\beta}(t+1)\boldsymbol{\eta}(t-1)$, where the $2Jp\times p(\sum_{b\in\mathcal{B}}|\mathcal{N}_b|)$ matrix $\check{\mathbf{R}}_h^{\beta}(t+1):=\left[\left[\check{\mathbf{R}}_h^{\beta}(t+1)\right]_{11}^T\left[\check{\mathbf{R}}_h^{\beta}(t+1)\right]_{21}^T\right]^T$ is given by

$$\begin{aligned}\left[\check{\mathbf{R}}_h^{\beta}(t+1)\right]_{11}&:=\frac{3\mu^{1-\gamma}c_1}{2}\mathbf{P}_{\alpha}-\left(\mu^{1-\gamma}c_1\right)^2\mathbf{A}\mathbf{P}_{\alpha}\\&\quad-2\mu^{2-\gamma}c_1\mathbf{R}_h(t+1)\mathbf{P}_{\alpha}-\mu^{1-\gamma}c_1\mathbf{P}_{\beta}\\\left[\check{\mathbf{R}}_h^{\beta}(t+1)\right]_{21}&:=\mu^{1-3\gamma/2}c_1c_2\mathbf{A}\mathbf{P}_{\alpha}\\&\quad+\mu^{-\gamma/2}c_2(\mathbf{P}_{\beta}-\mathbf{P}_{\alpha}).\end{aligned}$$

The remaining step is to show that $\check{\mathbf{R}}_h^{\beta}(t+1)=\mathrm{bdiag}\left(\mathbf{I}_{Jp},\mu^{-\gamma/2}c_2\mathbf{A}\right)\mathbf{R}_h^{\beta}(t+1)$. If the latter holds, then i) we group the noise terms in (59), (60) and the second term in (61), ii) take out $\mathrm{bdiag}\left(\mathbf{I}_{Jp},\mu^{-\gamma/2}c_2\mathbf{A}\right)$ as a common factor, and iii) conclude that $\mathbf{y}(t+1)$ is given by (32), while $\mathbf{z}(t+1)$ is provided by (33).

To show that $\check{\mathbf{R}}_h^{\beta}(t+1)=\mathrm{bdiag}\left(\mathbf{I}_{Jp},\mu^{-\gamma/2}c_2\mathbf{A}\right)\mathbf{R}_h^{\beta}(t+1)$, it suffices to prove that there exists matrix $\mathbf{C}_R$ such that $\mathbf{A}\mathbf{C}_R=\mathbf{P}_{\beta}-\mathbf{P}_{\alpha}$. To this end, it can be shown that (details omitted due to space limitations) $\mathrm{nullspace}\left(\mathbf{P}_{\beta}^T-\mathbf{P}_{\alpha}^T\right)=\mathrm{nullspace}(\mathbf{A})=\mathrm{span}(\bar{\boldsymbol{\nu}}_i)$, where $\bar{\boldsymbol{\nu}}_i:=\left[\boldsymbol{\nu}_{p,i}^T\ldots\boldsymbol{\nu}_{p,i}^T\right]^T$, $i=1,\ldots,p$. Since $\mathbf{A}$ is symmetric, we have $\mathrm{nullspace}(\mathbf{A})\perp\mathrm{range}(\mathbf{A})$. As $\mathrm{nullspace}\left(\mathbf{P}_{\beta}^T-\mathbf{P}_{\alpha}^T\right)\perp\mathrm{range}(\mathbf{P}_{\beta}-\mathbf{P}_{\alpha})$, it follows that $\mathrm{range}(\mathbf{P}_{\beta}-\mathbf{P}_{\alpha})\subseteq\mathrm{range}(\mathbf{A})$, which further implies that we can find $\mathbf{C}_R$ such that $\mathbf{A}\mathbf{C}_R=\mathbf{P}_{\beta}-\mathbf{P}_{\alpha}$. □

E. Proof of Lemma 4

We will specify first the step-size values for which $\mathbf{\Phi}(\mu):=E_{\mathbf{h}}[\mathbf{\Phi}(t+1,\mu)]$ has its eigenvalues inside the unit circle.

Lemma 6: *If c_1 is selected such that $c_1\lambda^{+}(\mathbf{A})>4$ and $\mu\in(0,\mu_u)$, where μ_u is defined in (38), then all the eigenvalues of $\mathbf{\Phi}(\mu)$ lie inside the unit circle, i.e., $|\lambda_i(\mathbf{\Phi}(\mu))|<1$ for $i=1,\ldots,2Jp$.*

Proof: Following steps similar to those in [18, App. H], we express the eigenvalues as roots of a second-order polynomial to determine bounds on μ that ensure $|\lambda_i(\mathbf{\Phi}(\mu))| < 1$. Further, the spectral radius of $\lambda_{\max}(\mu) := \lambda_{\max}(\mathbf{\Phi}(\mu))$, can be expressed as $\lambda_{\max}(\mu) = 1 - \mu^{1-\gamma} b_1 + \mu b_2 < 1$, where b_1, b_2 are constants with $b_1 > 0$. □

Using Lemma 6, we can apply the results in [21, Sec. C6, p. 321] to infer that for $\mu \in [0, \mu_u)$ there exists a finite constant $0 < \kappa(\mu)$, such that

$$\|\mathbf{\Phi}_1^t(\mu)\| \leq \kappa(\mu)\lambda_{\max}^t(\mu) \tag{62}$$

where $\mathbf{\Phi}_1(\mu)$ and $\mathbf{\Phi}_2(\mu)$ denote the $Jp \times 2Jp$ upper and lower block matrices of $\mathbf{\Phi}(\mu)$ obtained by keeping the upper Jp or lower Jp rows of $\mathbf{\Phi}(\mu)$, respectively.

In order to upper bound $\|\mathbf{\Phi}(1:t,\mu)\|$ in (39) we will establish a recursive inequality for $\|\mathbf{\Phi}(1:t,\mu)\|$ and then apply the discrete Bellman–Gronwall lemma [21, p. 315]. To this end, rewrite $\mathbf{\Phi}(t,\mu)$ as $\mathbf{\Phi}(t,\mu) = \mathbf{\Phi}(\mu) + \tilde{\mathbf{\Phi}}(t,\mu)$, where $\tilde{\mathbf{\Phi}}(t,\mu) := \mathrm{bdiag}(2\mu(\mathbf{R}_h - \mathbf{R}_h(t)), \mathbf{0}_{Jp})$. Then

$$\begin{aligned}\mathbf{\Phi}(1:t,\mu) &= \mathbf{\Phi}(t,\mu)\mathbf{\Phi}(1:t-1,\mu)\\ &= \mathbf{\Phi}(\mu)\mathbf{\Phi}(1:t-1,\mu) + \tilde{\mathbf{\Phi}}(t,\mu)\mathbf{\Phi}(1:t-1,\mu) \quad (63)\\ &= \mathbf{\Phi}^t(\mu) + \sum_{\tau=1}^{t} \mathbf{\Phi}^{t-\tau}(\mu)\tilde{\mathbf{\Phi}}(\tau,\mu)\mathbf{\Phi}(1:\tau-1,\mu) \quad (64)\end{aligned}$$

where $\mathbf{\Phi}(1:0,\mu) = \mathbf{I}_{Jp}$. Next, let $\mathbf{\Phi}_1(1:t,\mu)$ and $\mathbf{\Phi}_2(1:t,\mu)$ denote the $Jp \times 2Jp$ upper and lower block matrices of $\mathbf{\Phi}(1:t,\mu)$, respectively. Taking norm on both sides of (64) leads to the recursive inequality

$$\begin{aligned}\|\mathbf{\Phi}_1(1:t,\mu)\| &\leq \left\|\mathbf{\Phi}_1^t(\mu)\right\| \\ &\quad + \sum_{\tau=1}^{t} \left\|\mathbf{\Phi}_1^{t-\tau}(\mu)\right\| \|\tilde{\mathbf{\Phi}}(\tau,\mu)\|\|\mathbf{\Phi}_1(1:\tau-1,\mu)\| \quad (65)\\ &\leq \kappa(\mu)\lambda_{\max}^t(\mu) \\ &\quad + \sum_{\tau=1}^{t} \kappa(\mu)\lambda_{\max}^{t-\tau}(\mu)\|\tilde{\mathbf{\Phi}}(\tau,\mu)\|\|\mathbf{\Phi}_1(1:\tau-1,\mu)\| \quad (66)\end{aligned}$$

where the second inequality is obtained after using (62). Then, multiplying both sides of (65) with $\lambda_{\max}^{-t}(\mu)$, and applying the discrete Bellman–Gronwall lemma, leads to the following *non-recursive* inequality:

$$\begin{aligned}\|\mathbf{\Phi}_1(1:t,\mu)\| &\leq \kappa(\mu)\lambda_{\max}^t(\mu)\prod_{\tau=1}^{t}\left(1+\lambda_{\max}^{-1}(\mu)\kappa(\mu)\|\tilde{\mathbf{\Phi}}(\tau,\mu)\|\right)\\ &= \kappa(\mu)\prod_{\tau=1}^{t}(\lambda_{\max}(\mu) + \kappa(\mu)\|\tilde{\mathbf{\Phi}}(\tau,\mu)\|). \quad (67)\end{aligned}$$

Raising both sides of (67) to the power of $1/t$ and applying the arithmetic-mean geometric-mean inequality for the product term we arrive at

$$\begin{aligned}&\|\mathbf{\Phi}_1(1:t,\mu)\|^{1/t}\\ &\quad \leq \kappa^{1/t}(\mu)\left(\lambda_{\max}(\mu) + \mu\kappa(\mu)t^{-1}\sum_{\tau=1}^{t}\|\tilde{\mathbf{\Phi}}(\tau)\|\right) \quad (68)\end{aligned}$$

where $\tilde{\mathbf{\Phi}}(\tau) = \mathrm{bdiag}(2(\mathbf{R}_h - \mathbf{R}_h(\tau)), \mathbf{0}_{Jp})$.

Note at this point that $\lim_{t\to\infty}(\kappa(\mu))^{1/t} = 1$, while from a3) the strong law of large numbers implies that $\lim_{t\to\infty} t^{-1}\sum_{\tau=1}^{t}\|\tilde{\mathbf{\Phi}}(\tau)\| = E_{\mathbf{h}}[\|\tilde{\mathbf{\Phi}}(t)\|] = 2m_{\tilde{h}}$ exists a.s., and is bounded. The latter limits when combined with (68) give

$$\begin{aligned}\limsup_{t\to\infty}\|\mathbf{\Phi}_1(1:t,\mu)\|^{1/t} &\leq \lambda_{\max}(\mu) + 2\mu\kappa(\mu)m_{\tilde{h}}\\ &= 1 - \mu^{1-\gamma}\left[b_1 - \mu^{\gamma}(b_2 + 2\kappa(\mu)m_{\tilde{h}})\right] := \lambda'(\mu) \quad (69)\end{aligned}$$

a.s. Now, given that $b_1 > 0$ we can always find a $\mu_p > 0$ such that $b_1 - \mu^{\gamma}(b_2 + 2\kappa(\mu)m_{\tilde{h}}) > 0$ and $\lambda'(\mu) < 1$ for all $\mu \in (0, \mu_p]$. Thus, for $\mu \in (0, \mu_u')$ with $\mu_u' := \min(\mu_u, \mu_p)$ we ensure that (69) is satisfied, while $\lambda'(\mu) < 1$.

Next, we show that for $\mu \in (0, \mu_u')$ matrix $\mathbf{\Phi}_2(1:t,\mu)$ also satisfies an inequality of the form given in (69). Given that $\mu \in (0, \mu_u')$ it follows from (69) that there exists $t_0 < \infty$ and positive, a.s. finite random variable R such that

$$\|\mathbf{\Phi}_1(1:t,\mu)\| \leq (\lambda'(\mu))^t R, \quad \text{for all } t \geq t_0. \tag{70}$$

Recall that for $\mu \in (0, \mu_u')$ matrix $\mathbf{\Phi}(\mu)$ is stable (cf. Lemma 6); thus, similar to (62) we have $\|\mathbf{\Phi}_2^t(\mu)\| \leq \kappa'(\mu)\lambda_{\max}^t(\mu)$, where $\kappa'(\mu)$ is a positive constant.

Focusing on the lower part $\mathbf{\Phi}_2(1:t,\mu)$ and taking norms in (64) we obtain

$$\begin{aligned}\|\mathbf{\Phi}_2(1:t,\mu)\| &\leq \kappa'(\mu)\lambda_{\max}^t(\mu) + S(1:t;t)\\ &= \kappa'(\mu)\lambda_{\max}^t(\mu) + S(1:t_0;t) + S(t_0+1:t;t) \quad (71)\end{aligned}$$

where $S(t_1:t_2;t) := \sum_{\tau=t_1}^{t_2}\kappa'(\mu)\lambda_{\max}^{t-\tau}(\mu)\|\tilde{\mathbf{\Phi}}(\tau,\mu)\|\|\mathbf{\Phi}_1(1:\tau-1,\mu)\|$. Next, we provide bounds for the terms in (71) and examine how they behave as $t \to \infty$. To this end, let $\lambda(\mu) := \max(\lambda_{\max}(\mu), \lambda'(\mu))$, and use (70) to upper bound the third term in (71) as

$$\begin{aligned}&S(t_0+1:t;t)\\ &\quad \leq \kappa'(\mu)\sum_{\tau=t_0+1}^{t}\lambda^{t-\tau}(\mu)\|\tilde{\mathbf{\Phi}}(\tau,\mu)\|\lambda^{\tau-1}(\mu)R \quad (72)\\ &\quad \leq \kappa'(\mu)R\lambda^{t-1}(\mu)(t-t_0)\left[(t-t_0)^{-1}\sum_{\tau=t_0+1}^{t}\|\tilde{\mathbf{\Phi}}(\tau,\mu)\|\right]. \quad (73)\end{aligned}$$

The quantity inside the square brackets in (73) converges a.s. to $2\mu m_{\tilde{h}} < \infty$. Further, $R < \infty$ a.s. and $\lim_{t\to\infty}\lambda^{t-1}(\mu)(t - t_0) = 0$ since $\lambda(\mu) < 1$. Given these properties, it follows readily that $S(t_0+1:t;t)$ converges to zero a.s. as $t \to \infty$. The second term in (71) can be rewritten as

$$\begin{aligned}S(1:t_0;t) &= \kappa'(\mu)\lambda_{\max}^t(\mu)\\ &\quad \times \left[\sum_{\tau=1}^{t_0}\lambda_{\max}^{-\tau}(\mu)\|\tilde{\mathbf{\Phi}}(\tau,\mu)\|\|\mathbf{\Phi}_1(1:\tau-1,\mu)\|\right]\\ &= \kappa'(\mu)\lambda_{\max}^t(\mu)R' \leq \kappa'(\mu)\lambda^t(\mu)R' \quad (74)\end{aligned}$$

where R' is the quantity within the square brackets. Next, it suffices to show that R' is finite a.s. Since $t_0 < \infty$ we have to show that each of the summands within R' is finite a.s. Observe first that $\lambda_{\max}^{-\tau}(\mu) < \infty$ for $\tau = 1, \ldots, t_0$. Also, since $E[\|\tilde{\mathbf{\Phi}}(t,\mu)\|] < \infty$ [cf. a3)], it follows that $\|\tilde{\mathbf{\Phi}}(t,\mu)\| < \infty$ a.s.

Next, notice that $\|\mathbf{\Phi}(1:\tau,\mu)\| \leq \prod_{i=1}^{\tau}\|\mathbf{\Phi}(i,\mu)\|$ and recall that $E\left[\mathbf{h}_j(t)\mathbf{h}_j^T(t)\right] < \infty$ [cf. a3)] which further implies that $\|\mathbf{R}_h(t)\| < \infty$ a.s. Thus, $\|\mathbf{\Phi}(i,\mu)\| < \infty$ a.s. and consequently $\|\mathbf{\Phi}(1:\tau,\mu)\| \leq \infty$ a.s., for $\tau = 1,\ldots,t_0$. Using the previous bounds we can upper bound the rhs in (71). Then, raising this upper bound and the left-hand side of (71) to the power of $1/t$ we have

$$\|\mathbf{\Phi}_2(1:t,\mu)\|^{1/t} \leq \left(\kappa'(\mu) + R'\kappa'(\mu) + R\kappa'(\mu)\lambda^{-1}(\mu) \times (t-t_0)\left[(t-t_0)^{-1}\sum_{\tau=t_0+1}^{t}\|\tilde{\mathbf{\Phi}}(\tau,\mu)\|\right]\right)^{1/t}\lambda(\mu).$$

Since $R, R' < \infty$ a.s., while the sample average within the square brackets converges to $2\mu m_{\tilde{h}}$ a.s., it follows readily that the quantity multiplying $\lambda(\mu)$ converges to one as $t \to \infty$. Thus, $\limsup_{t\to\infty}\|\mathbf{\Phi}_2(1:t,\mu)\|^{1/t} \leq \lambda(\mu)$ a.s., while $\lambda(\mu) < 1$ for all $\mu \in (0,\mu_u')$. Combining the latter result with the one in (69), we deduce that $\limsup_{t\to\infty}\|\mathbf{\Phi}(1:t,\mu)\|^{1/t} \leq \lambda(\mu) < 1$, and consequently $\limsup_{t\to\infty}\log\left(\|\mathbf{\Phi}(1:t,\mu)\|^{1/t}\right) \leq \log(\lambda(\mu)) < 0$, a.s. □

F. Proof of Lemma 5

The proof follows readily from the result in [20, Sec. VI, (A13)]. Specifically, a3) implies that i) $\mathbf{\Phi}(t,\mu)$ is stationary and ergodic; ii) it holds that [cf. a3)–a4)]

$$E\left[\left\|\mathbf{R}_h^{\alpha}(t+1)\bar{\boldsymbol{\eta}}(t-1) + \mathbf{R}_h^{\beta}(t+1)\boldsymbol{\eta}(t-1) + [\boldsymbol{\epsilon}^T(t+1) \quad \mathbf{0}^T]^T\right\|\right] < \infty$$

since $E\left[\|\mathbf{R}_h^{\alpha}(t)\|\right] < \infty$, $E\left[\|\mathbf{R}_h^{\beta}(t)\|\right] < \infty$ and $E\left[\|\bar{\boldsymbol{\eta}}_j(t)\|\right] < \infty$; and iii) Lemma 4 shows that $\limsup_{t\to\infty} t^{-1}\log(\|\mathbf{\Phi}(1:t,\mu)\|) \leq \log(\lambda(\mu)) < 0$. Conditions i)–iii) guarantee that $\mathbf{z}(t)$ is weakly stochastically bounded [20, Sec. VI]. □

G. Proof of Proposition 1

Taking norms on both sides of (32) yields

$$\|\mathbf{y}(t+1)\| \leq \|\mathrm{bdiag}\left(\mathbf{I}_{Jp}, \mu^{-\gamma/2}c_2\mathbf{A}\right)\|\|\mathbf{z}(t+1)\| + \left\|\left[\mu^{1-\gamma}c_1\mathbf{I}_{Jp}^T - \mu^{-\frac{\gamma}{2}}c_2\mathbf{I}_{Jp}^T\right]^T\right\|(\|\bar{\boldsymbol{\eta}}(t)\| + \|\boldsymbol{\eta}^{\alpha}(t)\|). \tag{75}$$

For brevity, let $\xi(t)$ denote the sum of the last two terms in (75). Now recall that if Y_1, Y_2 are random variables and $Y_1 \leq Y_2$, then $\Pr[Y_1 \geq \delta] \leq \Pr[Y_2 \geq \delta]$; hence

$$\Pr[\|\mathbf{y}(t+1)\| \geq \delta] \leq \Pr\left[\left\|\mathrm{bdiag}\left(\mathbf{I}_{Jp}, \mu^{-\gamma/2}c_2\mathbf{A}\right)\right\| \times \|\mathbf{z}(t+1)\| + \xi(t) \geq \delta\right]. \tag{76}$$

Another property needed in the remainder of the proof is that if Y_1, Y_2 are positive random variables, then $\Pr[Y_1+Y_2 \geq \delta] \leq \Pr[Y_1 \geq \delta/2] + \Pr[Y_2 \geq \delta/2]$. Applying this property to the rhs of (76) yields

$$\Pr[\|\mathbf{y}(t+1)\| \geq \delta] \leq \Pr\left[\left\|\mathrm{bdiag}\left(\mathbf{I}_{Jp}, \mu^{-\gamma/2}c_2\mathbf{A}\right)\right\| \times \|\mathbf{z}(t+1)\| \geq \frac{\delta}{2}\right] + \Pr\left[\xi(t) \geq \frac{\delta}{2}\right].$$

Markov's inequality can now be used to obtain the upper bound $\Pr[\xi(t) \geq \delta/2] \leq 2\delta^{-1}E[\xi(t)]$. Boundedness of $E\left[\|\mathbf{R}_h^{\alpha}(t)\|\right]$, $E\left[\|\mathbf{R}_h^{\beta}(t)\|\right]$, $E[\|\bar{\boldsymbol{\eta}}(t)\|]$ and $\mu^{-\gamma/2}$, since $\mu \in (0,\mu_u')$ along with a4) imply that $2E[\xi(t)] = c_{\xi} < \infty$. Thus

$$\sup_{t\geq 0}\Pr[\|\mathbf{y}(t+1)\|\delta] \leq \sup_{t\geq 0}\Pr\Big[\|\mathbf{z}(t+1)\| \geq \left(\frac{\delta}{2}\right)\left\|\mathrm{bdiag}\left(\mathbf{I}_{Jp}, \mu^{-\gamma/2}c_2\mathbf{A}\right)\right\|^{-1}\Big] + c_{\xi}\delta^{-1}. \tag{77}$$

But from Lemma 5 we have that

$$\lim_{\delta\to\infty}\sup_{t\geq 0}\Pr\left[\|\mathbf{z}(t+1)\| \geq \delta/2 \times \left\|\mathrm{bdiag}\left(\mathbf{I}_{Jp}, \mu^{-\gamma/2}c_2\mathbf{A}\right)\right\|^{-1}\right] = 0$$

for any $\mu \in (0,\mu_u')$. Hence, letting $\delta \to \infty$ on both sides of (77) we find that $\mathbf{y}(t)$ is WSB; i.e.,

$$\lim_{\delta\to\infty}\sup_{t\geq 0}\Pr[\|\mathbf{y}(t+1)\| \geq \delta] = 0.$$

□

H. Proof of Proposition 2

Since $\mathbf{y}_1(t+1) - \bar{\mathbf{y}}_1(t+1) = (\mathbf{z}_1(t+1) - \bar{\mathbf{z}}_1(t+1))$, it suffices to show that

$$\lim_{\mu\to 0}\sup_{0\leq t\leq T/\mu^{1-\gamma-\beta}}\Pr[\|\bar{\mathbf{z}}_1(t) - \mathbf{z}_1(t)\| \geq \delta] = 0.$$

To this end, subtract $\bar{\mathbf{z}}_1(t+1)$ from $\mathbf{z}_1(t+1)$ and recursively substitute $\mathbf{z}_1(t) - \bar{\mathbf{z}}_1(t)$ to obtain

$$\begin{aligned}\mathbf{z}_1(t+1) - \bar{\mathbf{z}}_1(t+1) &= [[\mathbf{\Phi}(1:t+1,\mu)]_{11} \quad \mathbf{0}](\mathbf{z}(0) - \bar{\mathbf{z}}(0)) \\ &+ 2\mu\sum_{\tau=0}^{t}[[\mathbf{\Phi}(\tau+2:t+1,\mu)]_{11} \quad \mathbf{0}] \\ &\times \mathrm{diag}(\mathbf{R}_h(\tau+1) - \mathbf{R}_h, \mathbf{0})\bar{\mathbf{z}}(\tau) \\ &+ 2c_1\mu^{2-\gamma}\sum_{\tau=0}^{t}[\mathbf{\Phi}(\tau+2:t+1,\mu)]_{11} \\ &\times (\mathbf{R}_h - \mathbf{R}_h(\tau+1))\bar{\boldsymbol{\eta}}(\tau-1) \\ &+ 2c_1\mu^{2-\gamma}\sum_{\tau=0}^{t}[\mathbf{\Phi}(\tau+2:t+1,\mu)]_{11} \\ &\times (\mathbf{R}_h - \mathbf{R}_h(\tau+1))\mathbf{P}_{\alpha}\boldsymbol{\eta}(\tau-1)\end{aligned} \tag{78}$$

where $[\mathbf{\Phi}(\tau+2:t+1,\mu)]_{11}$ is the $Jp \times Jp$ upper left submatrix contained in $\mathbf{\Phi}(\tau+2:t+1,\mu)$. Now, let $T_1(0,t)$, $T_2(0,t)$, $T_3(0,t)$ denote the norm of each of the last three summands in (78), and $T(0,t) := \sum_{i=1}^{3}T_i(0,t)$. Since $\mathbf{z}(0) = \bar{\mathbf{z}}(0)$, it holds that $\|\mathbf{z}_1(t+1) - \bar{\mathbf{z}}_1(t+1)\| \leq T(0,t-t_0) + T(t-t_0+1,t)$, from which it follows that

$$\Pr[\|\mathbf{z}_1(t+1) - \bar{\mathbf{z}}_1(t+1)\| \geq \delta] \leq \Pr\left[T(0,t-t_0) \geq \frac{\delta}{2}\right] + \Pr\left[T(t-t_0+1,t) \geq \frac{\delta}{2}\right] \tag{79}$$

where $t_0 < \infty$ will be selected appropriately later on.

Thus, it suffices to prove that the two terms in (79) converge to zero as $\mu \to 0$. Toward this objective, the first term in (79) can be upper bounded as

$$\Pr\left[T(0,t-t_0) \geq \frac{\delta}{2}\right] \leq \Pr\left[T_1(0,t-t_0) \geq \frac{\delta}{4}\right] + \Pr\left[T_2(0,t-t_0) \geq \frac{\delta}{8}\right] + \Pr\left[T_3(0,t-t_0) \geq \frac{\delta}{8}\right]. \quad (80)$$

In order to show that the rhs in (80) goes to zero, recall from (62) that $\|[\mathbf{\Phi}^t(\mu)]_{11}\| \leq \kappa(\mu)\lambda_{\max}^t(\mu)$, where $\lambda_{\max}(\mu) = 1 - \mu^{1-\gamma}b_1 + \mu b_2$, with $b_1 > 0$ and $\kappa(\mu)$ finite positive constant. Note that a3) ensures that there exists finite t_0' such that $\left|(t_0')^{-1}\sum_{m=1}^{t_0'} \|\|\tilde{\mathbf{\Phi}}(m)\| - E[\|\tilde{\mathbf{\Phi}}\|]\right| \leq \Delta$ a.s. Then, setting $t_0 = t_0'$ and for $t \geq t_0'$, it follows from (68) that

$$\|[\mathbf{\Phi}(1:t,\mu)]_{11}\| \leq \kappa(\mu)\Big[\lambda_{\max}(\mu) + \mu\kappa(\mu)(E[\|\tilde{\mathbf{\Phi}}(t)\|] + \kappa_\Delta)\Big]^t \leq \kappa(\mu)\Big(1 - \mu^{1-\gamma}\Big(b_1 - \mu^\gamma(b_2 + \kappa(\mu)E[\|\tilde{\mathbf{\Phi}}\|] + \kappa(\mu)\kappa_\Delta)\Big)\Big)^t \quad (81)$$

where $\kappa_\Delta \in [-\Delta, \Delta]$. We now contend that $\sup_{0\leq t\leq T/\mu^{1-\gamma-\beta}} \Pr[T_1(0,t-t_0) \geq \delta/4] \to 0$ as $\mu \to 0$. Indeed, $T_1(0,t-t_0)$ can be bounded as [cf. (45)–(46)]

$$\begin{aligned} T_1(0,t-t_0) &\leq 2\mu \sum_{\tau=0}^{t-t_0} \|[\mathbf{\Phi}(\tau+2:t+1,\mu)]_{11}\| \|\tilde{\mathbf{\Phi}}(\tau+1)\| \\ &\quad \times \|[[\mathbf{\Phi}^\tau(\mu)]_{11}\,[\mathbf{\Phi}^\tau(\mu)]_{12}]\|\,\|\bar{\mathbf{z}}(0)\| \\ &\quad + 2\mu^{1+\beta} \sum_{\tau=0}^{t-t_0} \|[\mathbf{\Phi}(\tau+2:t+1,\mu)]_{11}\| \|\tilde{\mathbf{\Phi}}(\tau+1)\| \\ &\quad \times (\|\boldsymbol{\zeta}_1(\tau)\| + \|\boldsymbol{\zeta}_2(\tau)\| + \|\boldsymbol{\zeta}_3(\tau)\|) \end{aligned} \quad (82)$$

where

$$\begin{aligned} \boldsymbol{\zeta}_1(\tau) &:= \mu^{-\beta} \sum_{m=0}^{\tau-1} [[\mathbf{\Phi}^m(\mu)]_{11}\,[\mathbf{\Phi}^m(\mu)]_{12}]\,\mathbf{R}_h^\alpha \bar{\boldsymbol{\eta}}(\tau-m-2) \\ \boldsymbol{\zeta}_2(\tau) &:= \mu^{-\beta} \sum_{m=0}^{\tau-1} [[\mathbf{\Phi}^m(\mu)]_{11}\,[\mathbf{\Phi}^m(\mu)]_{12}]\,\mathbf{R}_h^\beta \boldsymbol{\eta}(\tau-m-2), \\ \boldsymbol{\zeta}_3(\tau) &:= \mu^{-\beta} \sum_{m=0}^{\tau-1} [\mathbf{\Phi}^m(\mu)]_{11}\,\boldsymbol{\epsilon}(\tau-m). \end{aligned} \quad (83)$$

Next, recall that $\|[\mathbf{\Phi}^m(\mu)]_{1i}\| \leq \kappa(\mu)\lambda_{\max}^t(\mu)$ with $\lambda_{\max}(\mu) = 1 - \mu^{1-\gamma}b_1 + \mu b_2$ and $\kappa(\mu) > 0$. Upon considering the expected norms of the noise terms in (83) and using a4), we arrive after tedious but straightforward manipulations at

$$\sup_{0\leq\tau\leq T/\mu^{1-\gamma-\beta}} E[\|\tilde{\mathbf{\Phi}}(\tau+1)\|\|\boldsymbol{\zeta}_i(\tau)\|] = m_{\zeta_i} < \infty, \quad i = 1,2,3. \quad (84)$$

Using the result in (84) in conjunction with (81) leads to

$$\lim_{\mu\to 0} \sup_{0\leq t\leq T/\mu^{1-\gamma-\beta}} E\,[T_1(0,t-t_0)] = 0. \quad (85)$$

Applying once more Markov's inequality for $\Pr(T_1(0,t-t_0) \geq \delta/4)$ yields

$$\lim_{\mu\to 0} \sup_{0\leq t\leq T/\mu^{1-\gamma-\beta}} \Pr\left[T_1(0,t-t_0) \geq \frac{\delta}{4}\right] = 0. \quad (86)$$

Using similar arguments we can bound the second and third probability terms in (80). Since $\sup_{0\leq\tau\leq T/\mu^{1-\gamma-\beta}} E\left[\|\tilde{\mathbf{\Phi}}(\tau+1)\|\|\bar{\boldsymbol{\eta}}_b(\tau-1)\|\right] < \infty$ and $\sup_{0\leq\tau\leq T/\mu^{1-\gamma-\beta}} E\left[\|\tilde{\mathbf{\Phi}}(\tau+1)\|\|\boldsymbol{\eta}(\tau-1)\|\right] < \infty$, use of (81) to bound $\|\mathbf{\Phi}(\tau+2:t+1,\mu)\|$, as well as Markov's inequality (as in $T_1(0,t-t_0)$) implies that $\lim_{\mu\to 0}\sup_{0\leq t\leq T/\mu^{1-\gamma-\beta}} \Pr[T_i(0,t-t_0) \geq \delta/8] = 0$, for $i = 2,3$. Combining these limits with the one in (86) establishes that

$$\lim_{\mu\to 0} \sup_{0\leq t\leq T/\mu^{1-\gamma-\beta}} \Pr\left[T(0,t-t_0) \geq \frac{\delta}{2}\right] = 0. \quad (87)$$

Consider next, the second probability term in (79). The three summands in (78) comprising $T_1(t-t_0+1,t)$ contain a finite number of terms, namely t_0. Hence, Markov's inequality yields $\Pr[T(t-t_0+1,t) \geq \delta/2] \leq (2/\delta)E[T(t-t_0+1,t)]$. Boundedness of the regressor moments further ensures that the expectation in the rhs of the last inequality converges to zero as $\mu \to 0$. We have already shown that the supremum of the two probability terms in (79) goes to zero as $\mu \to 0$ over the time interval $[0, T/\mu^{1-\gamma-\beta}]$; thus, the supremum of the left-hand side in (79) also goes to zero for vanishing μ. □

I. Proof of (48)

The covariance in (48) follows readily from (46) after recalling that $\boldsymbol{\eta}^\alpha(t) := \mathbf{P}_\alpha \boldsymbol{\eta}(t)$. Similarly, it is possible to find the covariance matrix of $\bar{\mathbf{z}}(t)$ in (49) using the recursive formula for $\bar{\mathbf{z}}(t)$ in (45) and setting $\breve{\boldsymbol{\eta}}(t+1) := \mathbf{R}_h^\alpha \bar{\boldsymbol{\eta}}(t-1) + \mathbf{R}_h^\beta \boldsymbol{\eta}(t-1) + [\boldsymbol{\epsilon}^T(t+1)\mathbf{0}^T]^T$. Thus, we can readily obtain $\mathbf{R}_{\breve{\eta}}$. Next, focus on the structure of $\mathbf{R}_{\bar{\eta}}$ and $\mathbf{R}_\eta$. From the definition in (24) it follows that $\mathbf{R}_{\bar{\eta}}$ consists of $p \times p$ submatrices of the form

$$[\mathbf{R}_{\bar{\eta}}]_{jj'} = \begin{cases} \sum_{b\in\mathcal{B}_j} \dfrac{\sum_{j''\in\mathcal{N}_b\setminus\{b\}} \mathbf{R}_{\eta_{b,j''}}}{|\mathcal{N}_b|^2} & \text{if } j' = j \text{ and } j \in \mathcal{J} \\ \sum_{b\in\mathcal{B}_j\cap\mathcal{B}_{j'}} \dfrac{\sum_{j''\in\mathcal{N}_b\setminus\{b\}} \mathbf{R}_{\eta_{b,j''}}}{|\mathcal{N}_b|^2} & \text{if } j' \neq j \text{ and } j \in \mathcal{J} \end{cases}$$

where $\mathbf{R}_{\bar{\eta}_{b,j}} := E\left[\bar{\boldsymbol{\eta}}_b^j(t)\left(\bar{\boldsymbol{\eta}}_b^j(t)\right)^T\right]$ denotes the covariance of the channel noise at bridge sensor b when receiving from sensor j. In the same way it follows from (31) that $\mathbf{R}_\eta$ is a block diagonal matrix with $(\sum_{b\in\mathcal{B}}|\mathcal{N}_b|)$ diagonal blocks of size $p\times p$. Each of these blocks is set equal to $\mathbf{R}_{\eta_{j,b}}$ of the corresponding channel noise vector $\boldsymbol{\eta}_j^b(t)$. Note also that $\mathbf{R}_{\eta_{j,j}} = \mathbf{0}$. □

REFERENCES

[1] Y. Bar-Shalom, X. R. Li, and T. Kirubarajan, *Estimation With Applications to Tracking and Navigation*. New York: Wiley-Interscience, 2001.

[2] D. P. Bertsekas and J. N. Tsitsiklis, *Parallel and Distributed Computation: Numerical Methods*, 2nd ed. Belmont, MA: Athena-Scientific, 1999.

[3] G. F. Franklin, J. D. Powell, and A. Emami-Naeini, *Feedback Control of Dynamic Systems*, 2nd ed. Reading, MA: Addison-Wesley, 1991.
[4] P. Gupta and P. R. Kumar, "The capacity of wireless networks," *IEEE Trans. Inf. Theory*, vol. 46, pp. 388–404, Mar. 2000.
[5] A. Jadbabaie and S. Morse, "Coordination of groups of mobile autonomous agents using nearest neighbor rules," *IEEE Trans. Autom. Control*, vol. 48, pp. 988–1001, Jun. 2003.
[6] O. Jahromi and P. Aarabi, "Distributed spectrum estimation in sensor networks," in *Proc. Int. Conf. Acoust., Speech, Signal Processing*, Montreal, QC, Canada, May 2004, pp. 849–852.
[7] H. J. Kushner and G. G. Yin, *Stochastic Approximation and Recursive Algorithms and Applications*, 2nd ed. New York: Springer-Verlag, 2003.
[8] C. G. Lopes and A. H. Sayed, "Incremental adaptive strategies over distributed networks," *IEEE Trans. Signal Process.*, vol. 55, no. 8, pp. 4064–4077, Aug. 2007.
[9] C. G. Lopes and A. H. Sayed, "Diffusion least-mean squares over adaptive networks: Formulation and performance analysis," *IEEE Trans. Signal Process.*, vol. 56, no. 7, pp. 3122–3136, Jul. 2008.
[10] G. Mateos, I. D. Schizas, and G. B. Giannakis, "Consensus-based distributed least-mean square algorithm using wireless sensor networks," presented at the 45th Allerton Conf., Monticello, IL, Sep. 2007.
[11] R. Olfati-Saber, "Distributed Kalman filtering for sensor networks," in *Proc. 46th Conf. Decision Control*, New Orleans, LA, Dec. 2007, pp. 5492–5498.
[12] R. Olfati-Saber and R. Murray, "Consensus problems in networks of agents with switching topology and time-delays," *IEEE Trans. Autom. Control*, vol. 49, pp. 1520–1533, Sep. 2004.
[13] R. Olfati-Saber and J. S. Shamma, "Consensus filters for sensor networks and distributed sensor fusion," in *Proc. 44th Conf. Dec. and Eur. Control Conf.*, Seville, Spain, Dec. 2005, pp. 6698–6703.
[14] C. H. Papadimitriou, *Computational Complexity*. Reading, MA: Addison-Wesley, 1993.
[15] S. S. Ram, A. Nedic, and V. V. Veeravalli, "Stochastic incremental gradient descent for estimation in sensor networks," in *Proc. Asilomar Conf. Signals, Systems, Computers*, Pacific Grove, CA, 2007, pp. 582–586.
[16] A. H. Sayed, *Fundamentals of Adaptive Filtering*. New York: Wiley, 2003.
[17] I. D. Schizas, G. B. Giannakis, A. Ribeiro, and S. I. Roumeliotis, "Consensus in ad hoc WSNs with noisy links—Part II: Distributed estimation and smoothing of random signals," *IEEE Trans. Signal Process.*, vol. 56, no. 4, pp. 1650–1666, Apr. 2008.
[18] I. D. Schizas, A. Ribeiro, and G. B. Giannakis, "Consensus in ad hoc WSNs with noisy links—Part I: Distributed estimation of deterministic signals," *IEEE Trans. Signal Process.*, vol. 56, no. 1, pp. 350–364, Jan. 2008.
[19] V. Solo, "On the stability of slowly time-varying linear systems," *Math. Control. Signals*, vol. 7, pp. 331–350, 1994.
[20] V. Solo, "The stability of LMS," *IEEE Trans. Signal Process.*, vol. 45, no. 12, pp. 3017–3026, Dec. 1997.
[21] V. Solo and X. Kong, *Adaptive Signal Processing Algorithms: Stability and Performance*. Englewood Cliffs, NJ: Prentice-Hall, 1995.
[22] D. P. Spanos, R. Olfati-Saber, and R. M. Murray, "Dynamic consensus on mobile networks," presented at the 16th IFAC World Congr., Prague, Czech, Jul. 2005.
[23] P. Stoica and R. Moses, *Spectral Analysis of Signals*. Englewood Cliffs, NJ: Prentice-Hall, 2005.
[24] B. Sundararaman, U. Buy, and A. D. Kshemkalyani, "Clock synchronization for wireless sensor networks: A survey," *Ad Hoc Netw.*, vol. 3, pp. 281–323, May 2005.
[25] J. Wu and H. Li, "A dominating-set-based routing scheme in ad hoc wireless networks," *Telecommun. Syst. J.*, vol. 3, pp. 63–84, Sep. 2001.
[26] L. Xiao and S. Boyd, "Fast linear iterations for distributed averaging," *Syst. Control Lett.*, vol. 53, pp. 65–78, Sep. 2004.
[27] L. Xiao, S. Boyd, and S. Lall, "A space–time diffusion scheme for peer-to-peer least-squares estimation," in *Proc. Int. Conf. Information Processing in Sensor Networks*, Nashville, TN, 2006, pp. 168–176.

Ioannis D. Schizas (S'07) received the Diploma degree in computer engineering and informatics (with Hons.) from the University of Patras, Patras, Greece, in 2004 and the M.Sc. degree in electrical and computer engineering from the University of Minnesota, Minneapolis, in 2007.

Since August 2004, he has been working towards the Ph.D. degree with the Department of Electrical and Computer Engineering, University of Minnesota, Minneapolis. His research interests lie in the areas of communication theory, signal processing, and networking. His current research focuses on distributed signal processing with wireless ad hoc and sensor networks, and distributed compression and source coding.

Gonzalo Mateos (S'07) received the B.Sc. degree in electrical engineering from the Universidad de la Republica Oriental del Uruguay, Montevideo, Uruguay, in 2005. Since August 2006, he has been working towards the Ph.D. degree at the University of Minnesota, Minneapolis.

Since 2003, he has been an Assistant with the Department of Electrical Engineering, Universidad de la Republica Oriental del Uruguay. From 2004 to 2006, he worked as a Systems Engineer at Asea Brown Boveri (ABB), Uruguay. He is currently a Research Assistant with the Department of Electrical and Computer Engineering, University of Minnesota, Minneapolis. His research interests lie in the areas of communication theory, signal processing, and networking. His current research focuses on distributed signal processing for wireless ad hoc and sensor networks.

Georgios B. Giannakis (F'97) received the Diploma degree in electrical engineering from the National Technical University of Athens, Greece, 1981 and the M.Sc. degree in electrical engineering, the M.Sc. degree in mathematics, and the Ph.D. degree in electrical engineering, all from the University of Southern California (USC), Los Angeles, in 1983, 1986, and 1986.

Since 1999, he has been a Professor with the Electrical and Computer Engineering Department at the University of Minnesota, Minneapolis, where he now holds an ADC Chair in Wireless Telecommunications. His general interests span the areas of communications, networking, and statistical signal processing—subjects on which he has published more than 250 journal papers, 450 conference papers, two edited books, and two upcoming research monographs on *Space-Time Coding for Broadband Wireless Communications* (Wiley, 2006) and *Ultra-Wideband Wireless Communications* (Cambridge Press, 2007). Current research focuses on diversity techniques, complex-field and space–time coding, multicarrier, cooperative wireless communications, cognitive radios, cross-layer designs, mobile ad hoc networks, and wireless sensor networks.

Dr. Giannakis is the (co)recipient of six paper awards from the IEEE Signal Processing (SP) and Communications Societies, including the G. Marconi Prize Paper Award in Wireless Communications. He also received Technical Achievement Awards from the SP Society (2000), from EURASIP (2005), a Young Faculty Teaching Award, and the G. W. Taylor Award for Distinguished Research from the University of Minnesota. He has served the IEEE in a number of posts.