STABILITY ANALYSIS OF THE CONSENSUS-BASED DISTRIBUTED LMS ALGORITHM

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

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

ABSTRACT

We deal with consensus-based online estimation and tracking of (non-) stationary signals using ad hoc wireless sensor networks (WSNs). A *distributed* (D-) least-mean square (LMS) like algorithm is developed, which offers simplicity and flexibility, while it solely relies on single-hop communications among sensors. Starting from a pertinent squared-error cost, we apply the alternating-direction method of multipliers to minimize it in a distributed fashion; and utilize stochastic approximation tools to eliminate the need for a complete statistical characterization of the processes of interest. By resorting to stochastic averaging and perturbed Lyapunov techniques, we further establish that local estimates are exponentially convergent to the true parameter of interest when observations are noise free and linearly related to it. This convergence result is necessary for bounding the estimation error in the presence of noise, and holds not only when regressors are white across time but even when they exhibit *temporal correlations*. Numerical tests confirm the merits of the novel D-LMS algorithm and its stability analysis.

Index Terms— Distributed estimation, Distributed algorithms, Adaptive signal processing

1. INTRODUCTION

As wireless sensor networks (WSNs) have become an emerging technology, distributed and collaborative estimation and tracking has drawn a lot of interest recently. Such tasks become more challenging under the severe communication constraints that sensors have to operate with. Previous works have mainly addressed (i) *single-shot* distributed estimation algorithms utilizing a snapshot of data, and (ii) distributed model-based Kalman filtering and smoothing. For related works on estimation using ad hoc WSNs, see e.g., [1] and references therein.

Here we deal with online estimation/tracking of (non-) stationary signals, incorporating new sensor data in real-time. Most importantly, we account for the fact that in many applications the environment experiences time variations (due to e.g., a changing WSN topology), and a complete statistical description of the underlying processes may not be available at the sensors. With similar concerns, distributed incremental strategies with embedded LMS-type adaptive filters at the sensors were introduced in [2]. Although [2] is attractive for its reduced communication burden, the requirement of a predefined connectivity cycle in the network that is also non-robust to sensor failures, poses an important limitation in large scale WSN deployments. Avoiding the need for such a cycle and further exploiting the exchange of information among single-hop neighbors to yield improved local estimates, the diffusion LMS [3] offers an improved alternative with increased communication cost and a somewhat heuristic derivation.

Recently, we developed a consensus-based distributed (D-) LMS algorithm for use in general ad hoc WSNs with noisy links, whose simplicity adheres to the power and communication resource scarcity characterizing these networks [4]. The algorithm is derived from a well-posed optimization problem defining the desired estimator, which we solve building on the techniques introduced in [1]. The main focus in [4] is on algorithmic aspects and the intuition behind the fusion of network-wide information to enrich local sensor estimates, as well as its inherent flexibility to trade-off communication cost for robustness to sensor failures. With respect to convergence, a first-order stability in the mean sense is reported under the widely assumed (though many times unrealistic) independence setting [5, Ch.5].

The main contribution of the present work is a stability result under significantly broadened signal assumptions, mainly uniform mixing of the regressor processes with mixing coefficients satisfying a summability condition. Drawing from stochastic averaging and perturbed Lyapunov techniques in [5, Ch.9], we establish that the local estimation errors associated with D-LMS are exponentially convergent to zero with probability 1, when the observations obey a noise-free linear model.

2. 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 the sensors in its neighborhood $\mathcal{N}_j \subseteq \{1, \ldots, J\} := \mathcal{J}$, where $j \in \mathcal{N}_j$. The communication links are assumed to be symmetric, and the WSN is modelled as an undirected graph whose vertices are the sensors and its edges represent the available links. Global connectivity information is summarized in the symmetric adjacency matrix $\mathbf{E} \in \mathbb{R}^{J \times J}$, where $[\mathbf{E}]_{ij} = 1$ if $i \in \mathcal{N}_j$ and 0 otherwise. As in [1], we assume that:

(a1) The communication graph is connected, meaning that there must exist a (possibly) multi-hop communication path connecting *any* two sensors. We will allow for non-ideal inter-sensor links (e.g., quantization, or, reception noise), whereby in general, a zero-mean additive noise vector denoted by $\mathbf{n}_j^i(t)$ will corrupt an information vector transmitted from sensor i to j at time instant t (t denotes discrete time). An example of such a network is shown in Fig. 1.

The WSN is deployed to estimate a signal vector $\mathbf{s}_0(t) \in \mathbb{R}^{p \times 1}$. For any given time instant t, each sensor has available a regressor vector $\mathbf{h}_j(t) \in \mathbb{R}^{p \times 1}$ and a scalar observation $x_j(t)$, which are assumed zero-mean without loss of generality. Further, we assume that:

(a2) The regressor vectors $\mathbf{h}_j(t)$ are wide sense stationary with covariance matrix $\bar{\mathbf{K}}_{h_j} = E[\mathbf{h}_j(t)\mathbf{h}_j^T(t)] > \mathbf{0}$ for $j = 1, \ldots, J$.

Upon defining the global quantities $\mathbf{x}(t) := [x_1(t), \ldots, x_J(t)]^T \in$

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

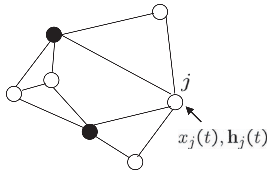

Fig. 1. An ad hoc WSN.

$\mathbb{R}^{J\times 1}$ and $\mathbf{H}(t) := [\mathbf{h}_1(t), \ldots, \mathbf{h}_J(t)]^T \in \mathbb{R}^{J\times p}$, the global least-mean squares estimation problem of interest can be posed as

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

Observe how the summands in the objective function are coupled through the global decision variable $\mathbf{s}$.

In the broad context of WSNs and targeting a low-complexity solution to (1), one could resort to a centralized (C-) LMS type of adaptation in a fusion center based topology. This, however, comes at the price of isolating the network's point of failure and increasing the communication cost (thus diminishing sensor battery lifetime) as the WSN scales. For these reasons, our goal is to develop and analyze a fully distributed LMS algorithm for use in ad hoc WSNs.

3. THE D-LMS ALGORITHM

In this section we present the D-LMS algorithm, first going through the process of algorithm construction and finally describing its operation. Our approach stems from three main building blocks: (i) recast (1) into an equivalent form amenable to distributed implementation, (ii) resort to the alternating-direction method of multipliers, see e.g., [1], so as to split the optimization problem into smaller subtasks that can be locally executed at the sensors, and (iii) use stochastic approximation tools to obtain an adaptive LMS-type of algorithm that can both handle the unavailability/variations of statistical information, and also be robust to change.

To this end, let us introduce the set of auxiliary variables $\boldsymbol{s} := \{\mathbf{s}_j\}_{j=1}^J$ that represent the local estimates at each of the sensors, and consider the convex *constrained* minimization problem

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

where $\mathcal{B} \subseteq \mathcal{J}$ is the bridge sensor subset introduced in [1]. The following conditions define a valid set $\mathcal{B}$: (i) $\forall\, j \in \mathcal{J}$ there exists at least one $b \in \mathcal{B}$ 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) if j_1 and j_2 are single-hop neighboring sensors, there must exist a bridge sensor b so that $b \in \mathcal{N}_{j_1} \cap \mathcal{N}_{j_2}$. A valid (not unique) bridge sensor assignment is shown in Fig. 1, where $\mathcal{B}$ consists of the sensors in black. An additional set of consensus-enforcing variables $\bar{\mathbf{s}} := \{\bar{\mathbf{s}}_b\}_{b\in\mathcal{B}}$ are maintained at each of the bridge sensors. It should be appreciated that the cost in (2) now has a decomposable structure whereas the constraints involve variables of neighboring sensors only. Interestingly, (a1) plus the defining characteristics of $\mathcal{B}$ provide necessary and sufficient conditions to assure that the equality constraints in (2) imply $\mathbf{s}_{j_1} = \mathbf{s}_{j_2}\ \forall\, j_1, j_2 \in \mathcal{J}$ [1]. This establishes the equivalence of (1) and (2) in the sense that $\hat{\mathbf{s}}_j(t) = \hat{\mathbf{s}}(t)\ \forall\, j \in \mathcal{J}$.

In order to solve (2), we consider its augmented Lagrangian given by

$$\mathcal{L}_a\left[\boldsymbol{s}, \bar{\mathbf{s}}, \mathbf{v}\right] = \sum_{j=1}^{J} E\left[\left(x_j(t+1) - \mathbf{h}_j^T(t+1)\mathbf{s}_j\right)^2\right]$$
$$+ \sum_{b\in\mathcal{B}} \sum_{j\in\mathcal{N}_b} \left[(\mathbf{v}_j^b)^T(\mathbf{s}_j - \bar{\mathbf{s}}_b) + \frac{c_j}{2}\|\mathbf{s}_j - \bar{\mathbf{s}}_b\|^2\right] (3)$$

where $\mathbf{v} := \left\{\mathbf{v}_j^b\right\}_{j\in\mathcal{J}}^{b\in\mathcal{B}_j}$ comprises the Lagrange multiplier vectors, and $c_j > 0$ are penalty coefficients introduced to tradeoff steady-state error for convergence speed. Application of the alternating-direction method of multipliers yields a three-step iterative update process which involves:

[S1] Updating multipliers $\mathbf{v}(t) = \{\mathbf{v}_j^b(t)\}_{j\in\mathcal{J}}^{b\in\mathcal{B}_j}$ via dual iterations.

[S2] For fixed $\bar{\mathbf{s}}(t) = \{\bar{\mathbf{s}}_b(t)\}_{j=1}^J$ and $\mathbf{v}(t)$ after completing **S1**, obtain $\boldsymbol{s}(t+1) = \{\mathbf{s}_j(t+1)\}_{j=1}^J$ as the minimizers of $\mathcal{L}_a[\boldsymbol{s}, \bar{\mathbf{s}}(t), \mathbf{v}(t)]$. In solving this convex and differentiable problem [cf. (3)], the first-order optimality condition yields an equation of the form $E[\mathbf{f}(\boldsymbol{s}, \mathbf{x}(t+1), \mathbf{H}(t+1))] = \mathbf{0}$ (details in [4]). In the absence of the (cross-) covariance information required to compute $\boldsymbol{s}(t+1)$, we propose an update recursion motivated by the Robbins-Monro algorithm (see e.g., [6, Ch.1]) using the noisy gradient $\mathbf{f}(\boldsymbol{s}, \mathbf{x}(t+1), \mathbf{H}(t+1))$.

[S3] Similarly, for fixed $\boldsymbol{s}(t+1)$ and $\mathbf{v}(t)$, obtain $\bar{\mathbf{s}}(t+1)$ as the minimizers of $\mathcal{L}_a\left[\boldsymbol{s}(t+1), \bar{\mathbf{s}}, \mathbf{v}(t)\right]$.

Remarkably, due to the structure of $\mathcal{L}_a\left[\boldsymbol{s}, \bar{\mathbf{s}}, \mathbf{v}\right]$ each of the computational tasks in **S2**-**S3** involves simpler minimization problems that can be locally tackled at the corresponding sensor.

For all sensors $j \in \mathcal{J}$, $b \in \mathcal{B}_j$ in (4) and $b \in \mathcal{B}$ in (6), the set of recursions (4)-(6) are the result of **S1**-**S3** and constitute the D-LMS algorithm with non-ideal communication links

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

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

$$\bar{\mathbf{s}}_b(t+1) = \sum_{j\in\mathcal{N}_b} \frac{\mathbf{v}_j^b(t) + c_j(\mathbf{s}_j(t+1) + \mathbf{n}_b^j(t+1))}{\sum_{r\in\mathcal{N}_b} c_r} \quad (6)$$

where in (5) $\mu_j > 0$ is a constant step-size allowing to track time-varying signals, and $e_j(t+1) := x_j(t+1) - \mathbf{h}_j(t+1)\mathbf{s}_j(t)$ is the local a priory error. We also note that the D-LMS recursions can be arbitrarily initialized.

The overall operation of the algorithm can be described as follows. At time instant t, sensor j receives the (noise corrupted) consensus variables $\bar{\mathbf{s}}_b(t) + \mathbf{n}_j^b(t)$ from its bridge neighbors $b \in \mathcal{B}_j$. Utilizing (4), it is able to update its Lagrange multipliers $\left\{\mathbf{v}_j^b(t)\right\}_{b\in\mathcal{B}_j}$ which are then used to compute $\mathbf{s}_j(t+1)$ via (5). Finally, sensor j transmits the quantity $c_j^{-1}\mathbf{v}_j^b(t) + \mathbf{s}_j(t+1)$ to all bridge sensors in its neighborhood $\mathcal{B}_j$. Consequently, each sensor $b \in \mathcal{B}$ acquires the vectors $\left\{c_j^{-1}\mathbf{v}_j^b(t) + \mathbf{s}_j(t+1) + \mathbf{n}_b^j(t+1)\right\}_{j\in\mathcal{N}_b}$ whose weighted average is computed using (6) to yield $\bar{\mathbf{s}}_b(t+1)$, thus completing the t-th iteration. Communication cost is $\mathcal{O}(p)$ per iteration, and only involves exchanges between neighboring sensors as required.

4. STABILITY ANALYSIS OF D-LMS

A widely adopted strategy in stability analysis of adaptive algorithms [5, Sec. 9.5] is to assume a linear observation model $x_j(t) = \mathbf{h}_j^T(t)\mathbf{s}_0 + \epsilon_j(t)$ per sensor and first neglect the observation noise. The results obtained from such an analysis should not be viewed as limited, since they can be used to ensure *boundedness* (in probability) of the D-LMS estimation error in the presence of observation and communication noise. Thus, in the sequel it is assumed that:

(a3) Sensor observations $x_j(t)$ obey the model $x_j(t) = \mathbf{h}_j^T(t)\mathbf{s}_0$. For simplicity in exposition we also select equal penalty coefficients $c_1 = \ldots = c_J = c$ and stepsizes $\mu_1 = \ldots = \mu_J = \mu$.

Under (a2) and two more assumptions on the regressor vectors, (to be specified later), we will establish that $\mathbf{s}_j(t) \to \mathbf{s}_0$ with probability 1 as $t \to \infty$. In order to facilitate the stability analysis of D-LMS we utilize (4)-(6) and apply simple algebraic manipulations to 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 $\mathbf{y}_{2,j}(t) := \sum_{b\in\mathcal{B}_j} \mathbf{v}_j^b(t)$ for $j \in \mathcal{J}$. Then, all the $\mathbf{y}_{1,j}(t)$ and $\mathbf{y}_{2,j}(t)$ vectors are stacked to form the supervectors $\mathbf{y}_1(t) := [\mathbf{y}_{1,1}^T(t)\ldots\mathbf{y}_{1,J}^T(t)]^T$ and $\mathbf{y}_2(t) := [\mathbf{y}_{2,1}^T(t)\ldots\mathbf{y}_{2,J}^T(t)]^T$ respectively, and $\mathbf{y}(t) := [\mathbf{y}_1^T(t)\ \ \mathbf{y}_2^T(t)]^T$. With $\otimes$ denoting Kronecker product, it turns out that D-LMS is equivalent to the following first-order linear updating rule (details in [7])

$$\mathbf{y}(t) = \mathbf{\Psi}(t,\mu)\mathbf{y}(t-1) \tag{7}$$

where for $t > 0$ the $2Jp \times 2Jp$ transition matrix $\mathbf{\Psi}(t,\mu)$ consists of four $Jp \times Jp$ matrix blocks given by $[\mathbf{\Psi}(t,\mu)]_{11} = \mathbf{I}_{Jp\times Jp} - 2\mu\mathbf{K}_h(t) - 2\mu c\mathbf{A}$, $[\mathbf{\Psi}(t,\mu)]_{12} = -\mu\mathbf{I}_{Jp\times Jp}$, $[\mathbf{\Psi}(t,\mu)]_{21} = c\mathbf{A}$ and $[\mathbf{\Psi}(t,\mu)]_{22} = \mathbf{I}_{Jp\times Jp}$, with

$$\mathbf{K}_h(t) := \text{bdiag}(\mathbf{h}_1(t)\mathbf{h}_1^T(t),\ldots,\mathbf{h}_J(t)\mathbf{h}_J^T(t)), \tag{8}$$

$$\mathbf{A} := \text{bdiag}(|\mathcal{B}_1|\mathbf{I}_{p\times p},\ldots,|\mathcal{B}_J|\mathbf{I}_{p\times p}) - \sum_{b\in\mathcal{B}} \frac{1}{|\mathcal{N}_b|}(\mathbf{e}_b\otimes\mathbf{I}_{p\times p})(\mathbf{e}_b\otimes\mathbf{I}_{p\times p})^T, \tag{9}$$

and $\mathbf{e}_b$ representing the b-th column of the adjacency matrix $\mathbf{E}$ while $\mathbf{I}_{p\times p}$ denoting the $p\times p$ identity matrix. For $t = 0$ the transition matrix in (7) takes the block-diagonal (bdiag) form $\mathbf{\Psi}(0,\mu) = \text{bdiag}(\mathbf{I}_{Jp\times Jp}, c\mathbf{A})$. Exploiting the structure of $\mathbf{\Psi}(t,\mu)$, recursion (7) can be rewritten as

$$\mathbf{y}(t) = \text{bdiag}(\mathbf{I}_{Jp\times Jp}, c\mathbf{A})\mathbf{z}(t) \tag{10}$$

where the state vector $\mathbf{z}(t)$ evolves as

$$\mathbf{z}(t) = \mathbf{\Phi}(t,\mu)\mathbf{z}(t-1) \tag{11}$$

while the transition matrix $\mathbf{\Phi}(t,\mu)$ consists of the submatrices $[\mathbf{\Phi}(t,\mu)]_{11} = \mathbf{I}_{Jp\times Jp} - 2\mu\mathbf{K}_h(t) - 2\mu c\mathbf{A}$, $[\mathbf{\Phi}(t,\mu)]_{12} = -\mu c\mathbf{A}$ and $[\mathbf{\Phi}(t,\mu)]_{21} = [\mathbf{\Phi}(t,\mu)]_{22} = \mathbf{A}\mathbf{A}^\dagger$, where † denotes pseudoinverse.

Careful inspection of (10) reveals that the local estimation errors in $\mathbf{y}_1(t)$ converge to zero as $t \to \infty$ as long as $\mathbf{z}_1(t)$ converges to zero. Thus, in order to ensure that $\lim_{t\to\infty}\mathbf{y}_1(t) = \mathbf{0}$ it suffices to analyze stability of the linear *time-varying stochastic* system in (11). To this end, we will rely on stochastic averaging and Lyapunov perturbation tools (see e.g., [5, Sect. 9.5]). First, consider an 'average' (in the stochastic sense) version of the system in (11) written as

$$\bar{\mathbf{z}}(t) = \bar{\mathbf{\Phi}}(\mu)\bar{\mathbf{z}}(t-1), \tag{12}$$

where $\bar{\mathbf{\Phi}}(\mu) := E[\mathbf{\Phi}(t,\mu)]$, while expectation is taken with respect to the regressors $\{\mathbf{h}_j(t)\}_{j=1}^J$. Note that the average system in (12) is linear *time-invariant* (LTI), and its transition matrix $\bar{\mathbf{\Phi}}(\mu)$ differs from $\mathbf{\Phi}(t,\mu)$ only in the upper left $Jp\times Jp$ matrix block. This block is written as $[\bar{\mathbf{\Phi}}(\mu)]_{11} = \mathbf{I}_{Jp\times Jp} - 2\mu\bar{\mathbf{K}}_h - 2\mu c\mathbf{A}$, where $\bar{\mathbf{K}}_h := E[\mathbf{K}_h(t)] = \text{bdiag}(\bar{\mathbf{K}}_{h_1},\ldots,\bar{\mathbf{K}}_{h_J})$. Due to the LTI structure of (12), its exponential stability to $\mathbf{0}$ boils down to satisfying $|\lambda_{\max}(\bar{\mathbf{\Phi}}(\mu))| < 1$, where $\lambda_{\max}(\cdot)$ denotes spectral radius. Using the eigen-decomposition of $\bar{\mathbf{\Phi}}(\mu)$, the latter inequality yields bounds for μ. Specifically, letting

$$\mu_u := 2\min\left(\lambda_{\max}^{-1}(\mathbf{K}_h + c\mathbf{A}), \lambda_{\max}^{-1}(2\mathbf{K}_h + \frac{3c}{2}\mathbf{A})\right),$$

we have established that [7]:

Lemma 1: *If $\mu \in (0,\mu_u)$ and $\lambda_{\max}(\mathbf{K}_h) < \infty$, then $\mathbf{0}$ is an exponentially stable equilibrium point for the average system in* (12)*; i.e., $|\lambda_{max}(\bar{\mathbf{\Phi}}(\mu))| < 1$ and for a finite constant $\alpha > 0$ it holds that*

$$\|\bar{\mathbf{z}}(t)\| \leq \alpha\lambda_{\max}^t(\bar{\mathbf{\Phi}}(\mu))\|\mathbf{z}(0)\|,\ \ \forall\ \ \mathbf{z}(0) \in \mathbf{R}^{2Jp\times 1}. \tag{13}$$

It is worth mentioning that the step-size bound μ_u is affected both by the regressors' covariance structure (as in the classical LMS) and also by the topology of the WSN through the matrix $\mathbf{A}$. Lemma 1 proves essential in showing that the D-LMS estimation error $\mathbf{y}_1(t)$ in (7) is exponentially convergent to $\mathbf{0}$. For that matter, we adopt the following assumptions:

(a4) Regressors $\mathbf{h}_j(t)$ obey a uniform mixing condition with mixing coefficients $\phi_{j,s} \geq 0$ for which $\sum_{s=0}^{\infty}\phi_{j,s}^{1/2} < \infty$. In detail, if $\mathcal{H}_j^{t_1,t_2}$ is the σ-algebra (history) generated by $\{\mathbf{h}_j(\tau)\}_{\tau=t_1}^{t_2}$, then for any events $H_1 \in \mathcal{H}_j^{0,t}$ and $H_2 \in \mathcal{H}_j^{t+s,\infty}$ it holds that

$$|\Pr(H_2|H_1) - \Pr(H_2)| \leq \phi_{j,s},\ \text{ with } \lim_{s\to\infty}\phi_{j,s} = 0. \tag{14}$$

(a5) The limit $\lim_{t\to\infty} t^{-1}\sum_{\tau=0}^{t}\|\mathbf{h}_j(\tau)\|_2^2$ exists and is finite.

Notice that (a4)-(a5) are quite general. In simple terms, these conditions impose the requirement that regressors $\mathbf{h}_j(t)$ and $\mathbf{h}_j(t+\tau)$ should become uncorrelated as $\tau \to \infty$. Thus, the stochastic process generating the regressors should have a decreasing temporal autocorrelation. This requirement is satisfied by many practical processes of interest, e.g., ARMA and is much more general than assuming that $\mathbf{h}_j(t)$ are white [3].

Next, the fact that the average system (12) is exponentially stable ensures existence of a Lyapunov function $V(\mathbf{z})$ for (12). In order to show that the estimation error associated with D-LMS goes to zero (exponential stability at zero), under (a1)-(a5), a *candidate* perturbed Lyapunov function (e.g., see [5, Sec. 9.5]) is formed for the system in (11) as

$$V(t,\mathbf{z}) = V(\mathbf{z}) + \mu\tilde{V}(t,\mathbf{z})$$

where $\tilde{V}(t,\mathbf{z})$ is the stochastic perturbation term designed to ensure that $V(t,\mathbf{z})$ is indeed a Lyapunov function for (11) when μ is sufficiently small (details in [7]). Interestingly, it can be shown that there exists $\mu_o \in (0,\mu_u)$ such that [7]:

Proposition 1: *Under (a1)-(a5) and if $\mu \in (0,\mu_o)$, then the D-LMS algorithm in* (7) *provides local estimates that are exponentially convergent to $\mathbf{s}_0$ with probability one (w.p. 1); i.e., for $j = 1,\ldots,J$*

$$\|\mathbf{y}_{1,j}(t)\| = \|\mathbf{s}_j(t) - \mathbf{s}_0\| \to \alpha\lambda^t(\mu),\quad w.p.1\ \text{ as } t\to\infty \tag{15}$$

where $\alpha > 0$ is a finite constant and $\lambda(\mu) < 1$.

Proposition 1 establishes almost sure convergence of D-LMS to the

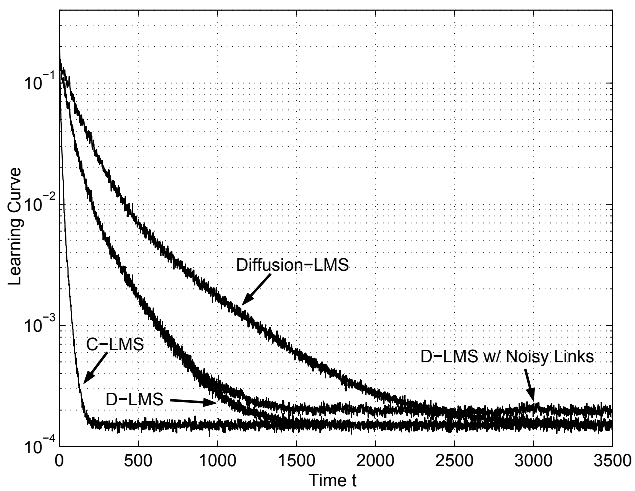

Fig. 2. Learning curve comparisons.

true parameter $\mathbf{s}_0$, across all sensors, for general temporally correlated regressors. This result goes beyond convergence in the mean sense [3, 4] and subsequently can be used to ensure that the estimation error associated with D-LMS is bounded in probability when: i) sensor data $x_j(t)$ contain observation noise; and ii) sensor communications are affected by additive noise (details in [7]).

5. SIMULATION RESULTS

Here we test the performance of D-LMS when the regressor vectors exhibit temporal correlation, comparing it with the diffusion LMS using Metropolis weights [3] and the C-LMS, which at every iteration uses all the available data in the network. We consider an ad hoc WSN with $J = 15$ sensors, obtained as a random geometric graph in $[0, 1] \times [0, 1]$ with communication range $r = 0.4$. The signal vector $\mathbf{s}_0$ has dimensionality $p = 2$, and for all $j = 1, \ldots, J$ the regressor vectors $\mathbf{h}_j(t) = [h_j(t), h_j(t-1)]^T$ have entries which evolve according to the large amplitude slowly time-varying process $h_j(t) = (1-\rho)\alpha_j h_j(t-1) + \sqrt{\rho}\nu_j(t)$. We have $\rho = 10^{-2}$, the $\alpha_j \sim \mathcal{U}[0, 1]$ (uniformly distributed) are i.i.d. and the driving white noise $\nu_j(t) \sim \mathcal{N}(0, \sigma^2_{\nu_j})$ (Gaussian) has a spatial variance profile given by $\sigma^2_{\nu_j} = 10^{-4}\beta_j$ with the $\beta_j \sim \mathcal{U}[0, 1]$ and i.i.d. A linear Gaussian model $\mathbf{x}(t) = \mathbf{H}(t)\mathbf{s}_0 + \epsilon(t)$ is adopted for the observations with $\sigma^2_{\epsilon_j} = 10^{-4}$. For all three algorithms we select $\mu = 10^{-2}$, and in particular for D-LMS $c_j = |\mathcal{B}_j|^{-1}$ for $j = 1, \ldots, J$. Finally, when testing D-LMS under noisy links we consider receiver AWGN with variance $\sigma^2_n = 10^{-4}$.

In Fig. 2 we compare the global mean-square error (MSE) evolution (learning curve) computed as $J^{-1}\sum_{j=1}^{J} E[\|\mathbf{x}(t) - \mathbf{H}(t)\mathbf{s}_j(t)\|^2]$ for the distributed approaches, where the average is taken over 60 realizations. As expected, C-LMS yields a performance benchmark while in all the (communication) noise-free cases, the resulting misadjustment is negligible. Furthermore, D-LMS outperforms the diffusion LMS and has a bounded MSE corroborating the stability result established in Proposition 1 which is necessary for convergence in the presence of observation noise [5, Sec. 9.6]. Further, the MSE remains bounded even when channel links are corrupted by reception noise, with increased steady-state MSE level as expected.

6. CONCLUSIONS

We have presented and analyzed the stability of a consensus-based distributed LMS algorithm suitable for operation in ad hoc WSNs. By resorting to the versatile stochastic averaging and perturbed Lyapunov analysis tools, we were able to establish that the D-LMS local estimates are almost surely exponentially convergent to the true parameters assuming a noise-free linear observation model, necessary for having estimation error boundedness in the presence of noise. These techniques are based on establishing a connection between the stability properties of the time-varying stochastic dynamical system of interest, with those of its (simpler) LTI averaged pair.

Numerical examples, involving correlated data, corroborated that the D-LMS outperforms existing online estimation schemes with comparable complexity. The algorithm is stable even in the presence of communication noise [7], whereas the steady-state MSE exhibits increased misadjustment as expected.

A D-RLS algorithm can also be obtained by directly applying the proposed approach to the exponentially weighted least-squares minimization problem. Interesting research directions emerge as we are naturally motivated to quantify analytically the convergence rate/estimation performance versus complexity tradeoff between these algorithms[1].

7. REFERENCES

[1] 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. on Signal Processing*, 2007, (to appear);http://spincom.ece.edu/.

[2] C. G. Lopes and A. H. Sayed, "Incremental adaptive strategies over distributed networks," *IEEE Trans. on Signal Processing*, vol. 55, pp. 4064–4077, August 2007.

[3] C. G. Lopes and A. H. Sayed, "Diffusion least-mean squares over adaptive networks," in *Proc. of the Intl. Conf. on Acoustics, Speech and Signal Proc.*, Honolulu, HI, Apr. 2007, vol. 3, pp. 509–512.

[4] G. Mateos, I. D. Schizas, and G. B. Giannakis, "Consensus-based distributed least-mean square algorithm using wireless sensor networks," in *Proc. of 45th Allerton Conference*, Monticello, IL, 2007.

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

[6] H. J. Kushner and G. G. Yin, *Stochastic Approximation and Recursive Algorithms and Applications*, Springer-Verlag, second edition, 2003.

[7] I. D. Schizas, G. Mateos, and G. B. Giannakis, "Distributed LMS for consensus-based in-network adaptive processing," *IEEE Trans. on Signal Processing*, Oct. 2007 (submitted).

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