Performance Analysis of Incremental LMS over Flat Fading Channels

Azam Khalili, Amir Rastegarnia

Abstract—We study the effect of fading in the communication channels between sensor nodes on the performance of the incremental least mean square (ILMS) algorithm, and derive steady state performance metrics, including the mean-square deviation (MSD), excess mean-square error (EMSE) and mean-square error (MSE). We obtain conditions for mean convergence of the ILMS algorithm, and show that in the presence of fading channels, the ILMS algorithm is asymptotically biased. Furthermore, the dynamic range for mean stability depends only on the mean channel gain, and under simplifying technical assumptions, we show that the MSD, EMSE and MSE are non-decreasing functions of the channel gain variances, with mean-square convergence to the steady states possible only if the channel gain variances are limited. We derive sufficient conditions to ensure mean-square convergence, and verify our results through simulations.

Index Terms—Adaptive networks, distributed estimation, incremental, least mean square.

I. INTRODUCTION

Many applications require us to learn or estimate some parameters related to a phenomenon of interest, based on all the available observations in a network [1]–[5]. Various schemes have been investigated in the literature to facilitate this, including the use of a fusion center to perform decentralized detection and estimation [6]–[8], and distributed methods that do not rely on a fusion center. An example of a distributed method is the consensus strategy [9]–[11] in which each node performs a local estimation and fuses its estimate with those of its neighbors so that all nodes converge to the same estimate as the number of iterations increases. An alternative distributed approach is the incremental update method [12]–[17], which relies on nodes passing updates to each other in a Hamiltonian cycle in the network. Yet another method is the diffusion strategy [18]–[25] that performs online estimation in a distributed manner by letting each node update its local estimate based on information from all its neighbors and over multiple observation epochs. In both the incremental update and diffusion approaches, techniques from distributed optimization [26], [27] are incorporated into the updates and local estimates need not converge to the same value, which leads to better performance over consensus strategies [28]. Networks that rely on in-network processing at each node while allowing the node estimates to update and adapt to new sensor observations have come to be known as adaptive networks [12], [22]–[24].

One of the most popular approaches to modeling the underlying process observed by the nodes in an adaptive network is to adopt a linear model and use a least mean squares (LMS) criterion in the estimation procedure. This is

because of its simplicity and wide applicability, for example in localization of targets, collaborative spectrum sensing, and in modeling group behaviors in biological systems [5], [24]. In this paper, we investigate the performance of the incremental LMS (ILMS) algorithm in a wireless adaptive network with communication links between neighboring nodes modeled as fading channels. In the original ILMS strategy proposed by [29], it is assumed that nodes communicate with each other via ideal links. However, this is typically not true in practice. In [30]–[34], the effect of additive link noise in the communication channels between nodes have been investigated. In [35], the performance of general adaptive diffusion algorithms in the presence of imperfect information exchanges, including quantization errors, and model non-stationarities has been considered. All these works however do not apply directly to a wireless sensor network, whose communication links are usually modeled by fading channels [36]. The references [37], [38] propose diffusion LMS algorithms for wireless sensor networks with fading channels but under the assumption that channel state information is known so that channel equalization can be performed. However, in practice the fading coefficients can only be estimated up to an uncertainty. Moreover, such channel equalization may be impractical for channels with time varying channel information. As wireless sensor networks have many important practical applications like building structure monitoring, it is therefore important to investigate the performance of adaptive networks in the presence of fading channels, without full knowledge of channel state information.

In this paper, our objective is to investigate how the performance of the ILMS algorithm is affected by the channel fading statistics of the communication channels between nodes. We restrict our analysis to the ILMS algorithm instead of the more general class of diffusion algorithms because of the significantly more complex analysis involved for diffusion algorithms, which is out of the scope of this paper. Nevertheless, to the best of our knowledge, this work is the first to analytically evaluate the performance of the ILMS algorithm in a practical wireless sensor network with fading channels, without exact knowledge of channel state information. We believe that many of the insights obtained in this paper are applicable to more general diffusion algorithms, which will be addressed as part of future work. Our main contributions in this paper are the following:

(i) We show that the ILMS algorithm over fading channels, like the traditional ILMS algorithm, is stable in mean if the step size is chosen in an appropriate range. We show

that this range now not only depends on the regression correlation, but also on the mean of the channel fading gains. Moreover, our analysis reveals that in general, fading communication channels lead to biased estimates in steady state.

(ii) We derive closed-form expressions for the steady state performance metrics, including the mean-square deviation (MSD), excess mean-square error (EMSE) and mean-square error (MSE), of the ILMS algorithm in the presence of fading channels, and under a Gaussian model. We show explicitly how these metrics are affected by the fading channel statistics.

(iii) We derive sufficient conditions for the convergence of the MSD, EMSE and MSE, and show that for a fixed step size, mean-square stability is lost if the channel gain variances become large. Under simplifying technical conditions, we show that under the decibel (dB) scale, the MSD, EMSE and MSE are approximately non-decreasing linear functions of the channel gain second order moments.

We also perform extensive simulations to verify that our theoretical analysis closely matches the actual steady state performance observed in a network.

The remainder of this paper is organized as follows. In Section II, we describe our system model and assumptions. In Section III, we present theoretical analysis of the steady state performance of the ILMS algorithm over fading channels, and discuss some of our results. In Section IV, we present simulation results to verify our theoretical analysis, and we conclude in Section V.

Notation: We adopt boldface letters for random quantities. The symbol $*$ denotes conjugation for scalars and Hermitian transpose for matrices. The notation $\mathrm{col}\{\cdot\}$ denotes a column vector (or matrix) with the specified entries stacked on top of each other. The notation $\mathrm{diag}\{\cdot\}$ will be used in two ways: $X = \mathrm{diag}\{x\}$ is a diagonal matrix whose entries are those of the vector x, and $x = \mathrm{diag}\{X\}$ is a vector containing the main diagonal of X. The exact meaning of this notation will be clear from the context. We let $\mathbb{1} \triangleq \mathrm{diag}\{I\}$ denote a vector consisting of all 1s. We use $\lambda_{\max}(A)$, $\lambda_{\min}(A)$ and $\rho(A)$ to denote the largest eigenvalue, smallest eigenvalue, and spectral radius of A, respectively. If Σ is a matrix, we use the notation $\|x\|_{\Sigma}^2 = x^*\Sigma x$ for the weighted square norm of x. If σ is a vector, the notation $\|x\|_{\sigma}$ is used to represent $\|x\|_{\mathrm{diag}\{\sigma\}}$.

II. PROBLEM FORMULATION

In this section, we first describe the data model, and give a brief overview of the traditional ILMS algorithm in [29]. We then describe how the ILMS updates are changed when information is communicated over fading channels. Finally, we list the assumptions we are making throughout this paper.

Consider a network composed of N nodes. At time i, node k observes a scalar measurement $\boldsymbol{d}_k(i)$ and a $1 \times M$ regression vector $\boldsymbol{u}_{k,i}$, which are related via a linear regression model

$$\boldsymbol{d}_k(i) = \boldsymbol{u}_{k,i} w^o + \boldsymbol{v}_k(i), \tag{1}$$

where $\boldsymbol{v}_k(i)$ is the observation (or measurement) noise, and the $M \times 1$ vector $w^o = [w^o(1), w^o(2), \cdots, w^o(M)]^T$ is a

deterministic (but unknown) vector. Based on the intended application, w^o may have different physical meanings. For example, w^o may represent the location of a target in space, or the parameters of an AR model [24], [25]. We make the following assumptions regarding the data model in (1). These assumptions are commonly assumed in the literature [24], [25], [29].

Assumption 1:

(i) For $k = 1, \ldots, N$ and $i \geq 1$, the regression vectors $\boldsymbol{u}_{k,i}$ are independent over node indices k and observation times i.

(ii) The measurement noises $\boldsymbol{v}_k(i)$ for all nodes $k = 1, \ldots, N$, and all observation times $i \geq 1$, are zero-mean and independent of each other and the regression vectors $\boldsymbol{u}_{k,i}$.

The goal of the network is to estimate w^o, at every node k, using all observed data in the entire network. Mathematically, the network seeks to find w^o that minimizes the following LMS objective function:

$$\sum_{k=1}^{N} \mathbb{E}\left[|\boldsymbol{d}_k(i) - \boldsymbol{u}_{k,i} w|^2\right]. \tag{2}$$

It can be shown that the minimizer w^o of (2) satisfies the normal equation [39].

$$w^o = \left(\sum_{k=1}^{N} R_{u,k}\right)^{-1} \left(\sum_{k=1}^{N} r_{du,k}\right),$$

where

$$R_{u,k} = \mathbb{E}[\boldsymbol{u}_{k,i}^* \boldsymbol{u}_{k,i}], \quad \text{and} \quad r_{du,k} = \mathbb{E}[\boldsymbol{d}_k(i)\boldsymbol{u}_{k,i}^*].$$

A. The ILMS Algorithm over Ideal Links

In [12], [29], an adaptive network based on incremental cooperation amongst nodes at every observation time has been developed wherein nodes communicate through a pre-established Hamilton cycle. This is known as the ILMS algorithm. Each node k receives a local estimate from the previous node $k-1$, updates it using its local data, and then sends it to the next node $k+1$, where node indices are modulo N. The update equations for the ILMS algorithm, at iteration i is given by [12]

$$\begin{cases} \boldsymbol{w}_{0,i} \leftarrow \boldsymbol{w}_{N,i-1} \\ \boldsymbol{w}_{k,i} = \boldsymbol{w}_{k-1,i} + \mu_k \boldsymbol{u}_{k,i}^* (\boldsymbol{d}_k(i) - \boldsymbol{u}_{k,i} \boldsymbol{w}_{k-1,i}) \end{cases} \tag{3}$$

where $\boldsymbol{w}_{k,i}$ is the local estimate of the node k at time i. It is shown in [12], [24] that as $i \to \infty$, we have $\boldsymbol{w}_{k,i} \to w^o$ in the mean, for every node k, and for an appropriately chosen set of step sizes $\{\mu_k : k = 1, \ldots, N\}$.

B. The ILMS Algorithm over Fading Channels

In this paper, we consider the case where nodes communicate over fading channels (see Fig. 1). To model model the Impact of channel, let denote by $\boldsymbol{r}_{k,i}$ the received signal at node k and time i. By incorporating the impact of fading channels we have

$$\boldsymbol{r}_{k,i} = \boldsymbol{h}_k(i)\boldsymbol{w}_{k-1,i} + \boldsymbol{q}_{k,i} \tag{4}$$
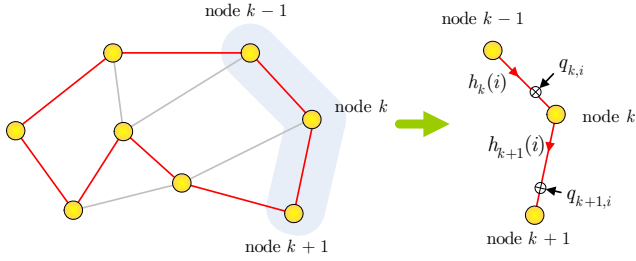
Fig. 1.   The incremental LMS adaptive network with fading channels.

Using (4), the update equation for node k in (3) becomes

$$\boldsymbol{w}_{k,i} = \boldsymbol{r}_{k,i} + \mu_k \boldsymbol{u}_{k,i}^* (\boldsymbol{d}_k(i) - \boldsymbol{u}_{k,i} \boldsymbol{r}_{k,i}) \qquad (5)$$

where $\boldsymbol{h}_k(i)$ is the channel gain at time i for the communication channel between node $k - 1$ and k, and $\boldsymbol{q}_{k,i}$ is the additive channel noise. The update equation (5) in terms of $\boldsymbol{h}_k(i)$, $\boldsymbol{w}_{k-1,i}$ and $\boldsymbol{q}_{k,i}$ as

$$\begin{aligned}
\boldsymbol{w}_{k,i} = {} & \boldsymbol{h}_k(i) \boldsymbol{w}_{k-1,i} + \boldsymbol{q}_{k,i} \\
& + \mu_k \boldsymbol{u}_{k,i}^* \left(\boldsymbol{d}_k(i) - \boldsymbol{u}_{k,i}(\boldsymbol{h}_k(i) \boldsymbol{w}_{k-1,i} + \boldsymbol{q}_{k,i}) \right), \quad (6)
\end{aligned}$$

We make the following assumptions regarding the fading channel statistics.

Assumption 2:

1) We assume phase coherent reception at every node k and model the channel gain as a non-negative random variable with $m_k = \mathbb{E}[\boldsymbol{h}_k(i)]$ and $s_k = \mathbb{E}\left[\boldsymbol{h}_k^2(i)\right]$ as the mean and second order moment of the channel gain for node k, respectively.

2) The channel gains $\boldsymbol{h}_k(i)$ for all nodes $k = 1, \ldots, N$, and all observation times $i \geq 1$, are independent of each other. For each node k, the channel gains $\{\boldsymbol{h}_k(i) : i \geq 1\}$ are identically distributed.

3) The channel gains $\boldsymbol{h}_k(i)$ for all nodes $k = 1, \ldots, N$, and all observation times $i \geq 1$, are independent of $(\boldsymbol{d}_l(j), \boldsymbol{u}_{l,j})$ for all l and j.

4) The additive channel noises $\boldsymbol{q}_{k,i}$ for all nodes $k = 1, \ldots, N$, and all observation times $i \geq 1$, are zero-mean and independent of each other, the observation noises $\boldsymbol{v}_l(j)$, and the regression vectors $\boldsymbol{u}_{l,i}$ for all l and j. The channel noise $\boldsymbol{q}_{k,i}$ has covariance matrix $Q_k = \mathbb{E}\left[\boldsymbol{q}_{k,i} \boldsymbol{q}_{k,i}^*\right]$.

Remark 1: When the channel fading coefficients $\boldsymbol{h}_k(i)$ are available, we can use zero-forcing (ZF) equalizer to mitigate the effects of fading channels. To show this, let denote by $\boldsymbol{z}_{k,i}$ the ZF equalizer coefficient which is given by

$$\boldsymbol{z}_k(i) = \frac{\boldsymbol{h}_k^*(i)}{\|\boldsymbol{h}_k(i)\|^2} \qquad (7)$$

Multiplying the received signal with the ZF equalizer coefficient gives

$$\boldsymbol{r}_{k,i}' = \boldsymbol{z}_k(i) \boldsymbol{r}_{k,i} = \boldsymbol{w}_{k-1,i} + \boldsymbol{q}_{k,i}' \qquad (8)$$

where $\boldsymbol{q}_{k,i}' \triangleq \frac{\boldsymbol{h}_k^*(i)}{\|\boldsymbol{h}_k(i)\|^2} \boldsymbol{q}_{k,i}$ can be interpreted as the effect of fading channel in the special case when the fading coefficients

$\boldsymbol{h}_k(i)$ are available. We can see that in this case the ILMS algorithm with fading channels behaves like the ILMS network with noisy links.

Remark 2: In practice, the channel fading coefficients $\boldsymbol{h}_k(i)$ are not available and only can be estimated by e.g. training data. In this case even if the nodes perform channel state estimation, it is not possible to measure the channel gains with absolute certainty, especially if the channels experience fast fading. To further the motivation for using (6), let $\hat{\boldsymbol{h}}_k(i)$ be the estimate of the fading coefficient $\boldsymbol{h}_k(i)$. Then, the ZF equalizer coefficient becomes

$$\boldsymbol{z}_k'(i) = \frac{\hat{\boldsymbol{h}}_k^*(i)}{\|\hat{\boldsymbol{h}}_k(i)\|^2} \qquad (9)$$

Using (9) $\boldsymbol{r}_{k,i}'$ in (8) changes to

$$\boldsymbol{r}_{k,i}' = \boldsymbol{z}_k'(i) \boldsymbol{r}_{k,i} = \boldsymbol{h}_k'(i) \boldsymbol{w}_{k-1,i} + \boldsymbol{q}_{k,i}'' \qquad (10)$$

where

$$\boldsymbol{h}_k'(i) \triangleq \frac{\boldsymbol{h}_k(i) \hat{\boldsymbol{h}}_k^*(i)}{\|\hat{\boldsymbol{h}}_k(i)\|^2}, \quad \boldsymbol{q}_{k,i}'' \triangleq \frac{\hat{\boldsymbol{h}}_k^*(i)}{\|\hat{\boldsymbol{h}}_k(i)\|^2} \boldsymbol{q}_{k,i} \qquad (11)$$

We can see from (10) that when the nodes perform channel state estimation the channel gain for node k can still be modeled as a non-negative random variable with a non-trivial variance. Of particular interest is the case where channel estimation is sufficiently accurate on average so that the channel gain can be partially accounted for at the receiver node. In this case, we can assume that $m_k = 1$ and $s_k > 1$. In the next section, we show how the performance of the ILMS algorithm depends on both m_k and s_k.

III. PERFORMANCE ANALYSIS

In this section, we analyze the mean stability and steady state mean-square performance of the ILMS algorithm when communication channels between nodes are fading channels. Our analysis is based on the energy conservation approach of [12], [39].

In our analysis, we will use the deviation between an observed measurement and its prediction based on the current local estimate, which is defined as

$$\boldsymbol{e}_k(i) = \boldsymbol{d}_k(i) - \boldsymbol{u}_{k,i} \boldsymbol{w}_{k-1,i},$$

and the weight error vector, which is the deviation between the local estimate $\boldsymbol{w}_{k,i}$ and its true value w^o, given by

$$\widetilde{\boldsymbol{w}}_{k,i} = w^o - \boldsymbol{w}_{k,i}.$$

It must be noted that in mean and mean-square analysis of the ILMS algorithm with fading channels, we need to evaluate the first and second order moments of the weight error vector $\widetilde{\boldsymbol{w}}_{k,i}$.

A. Mean Stability Analysis

Consider the update equation (6). By subtracting w^o from both sides of (6) and using the definition of $\widetilde{\boldsymbol{w}}_{k,i}$ we obtain

$$\begin{aligned}
\widetilde{\boldsymbol{w}}_{k,i} = {}& \boldsymbol{h}_k(i)\widetilde{\boldsymbol{w}}_{k-1,i} \\
& + (1 - \boldsymbol{h}_k(i))w^o - \mu_k \boldsymbol{h}_k(i)\boldsymbol{u}_{k,i}^* \boldsymbol{u}_{k,i}\widetilde{\boldsymbol{w}}_{k-1,i} \\
& - \mu_k \boldsymbol{u}_{k,i}^* \boldsymbol{v}_k(i) - \mu_k(1 - \boldsymbol{h}_k(i))\boldsymbol{u}_{k,i}^* \boldsymbol{u}_{k,i}w^o - \boldsymbol{q}_{k,i} \\
& + \mu_k \boldsymbol{u}_{k,i}^* \boldsymbol{u}_{k,i}\boldsymbol{q}_{k,i}.
\end{aligned} \tag{12}$$

The recursion (12) shows how the weight error vector $\widetilde{\boldsymbol{w}}_{k,i}$ evolves over each update of the ILMS algorithm at time i. We use it to derive the required sufficient conditions for mean convergence of the ILMS algorithm with fading channels in the following result.

Proposition 1: Under Assumptions 1 and 2, the ILMS algorithm over fading channels, given by the update equation (6), is stable in mean for any initial conditions if we have

$$\max\left\{0, \frac{m_k - 1}{m_k \lambda_{\min}(R_{u,k})}\right\} < \mu_k < \frac{m_k + 1}{m_k \lambda_{\max}(R_{u,k})}. \tag{13}$$

Proof: To derive the stability condition in the mean, we need to derive a condition on the step-size in order to guarantee convergence in the mean so that $\lim_{i\to\infty} \mathbb{E}[\widetilde{\boldsymbol{w}}_{k,i}] = 0$. To this end we consider again (12). By taking expectation on both sides of (12) and using Assumptions 2 yields

$$\mathbb{E}[\widetilde{\boldsymbol{w}}_{k,i}] = m_k J_k \mathbb{E}[\widetilde{\boldsymbol{w}}_{k-1,i}] + (1 - m_k)J_k w^o \tag{14}$$

with

$$J_k \triangleq I - \mu_k R_{u,k}. \tag{15}$$

By iterating (14), we obtain that the weight error vector evolves according to

$$\begin{aligned}
\mathbb{E}[\widetilde{\boldsymbol{w}}_{k,i}] = {}& \left(\prod_{n=1}^{N} m_n J_n\right) \mathbb{E}[\widetilde{\boldsymbol{w}}_{k,i-1}] \\
& + \sum_{n=1}^{N}\left((1 - m_n)J_n \prod_{\ell=n+1}^{N} m_\ell J_\ell\right)w^o.
\end{aligned} \tag{16}$$

We can see from (16) that convergence in mean for the ILMS algorithm with fading channels depends on the modes of the matrix

$$\mathcal{M} \triangleq \prod_{n=1}^{N} m_n J_n. \tag{17}$$

The necessary and sufficient condition required for convergence of (6) in the mean is that the matrix $\mathcal{M}$ be stable. Equivalently, all eigenvalues of $\mathcal{M}$ must be inside the unit circle, i.e.,

$$\rho(\mathcal{M}) < 1, \tag{18}$$

We already know that the spectral radius of any matrix X we have $\rho(X) < \|X\|$ for any induced matrix norm [40]. Thus we have

$$\begin{aligned}
\rho(\mathcal{M}) & \leq \|\mathcal{M}\| \\
& \leq \|m_1 J_1\| \|m_2 J_2\| \cdots \|m_N J_N\| \\
& = \rho(m_1 J_1)\rho(m_2 J_2) \cdots \rho(m_N J_N)
\end{aligned}$$

Since every $\{J_k\}$ is a Hermitian matrix, its 2-induced norm agrees with its spectral radius, which explains the last equality. Thus, to guarantee that for any k the constraint (18) is satisfied if for all $k = 1, \ldots, N$, it is enough to have $\rho(m_k J_k) \leq 1$ which can be stated in terms of the λ of $R_{u,k}$ as

$$|m_k(1 - \mu_k \lambda)| < 1. \tag{19}$$

Therefore, if the step-size μ_k is chosen such that

$$\max\left\{0, \frac{m_k - 1}{m_k \lambda_{\min}(R_{u,k})}\right\} < \mu_k < \frac{m_k + 1}{m_k \lambda_{\max}(R_{u,k})}$$

convergence of $\mathbb{E}[\widetilde{\boldsymbol{w}}_{k,i}]$ in the mean is guaranteed, and the proof is complete. ∎

Remark 3: It is worth noting that for the traditional ILMS algorithm in (3), where information is exchanged perfectly over ideal links (i.e., we have $m_k = 1$ for all nodes k), the convergence condition (13) reduces to

$$0 < \mu_k < \frac{2}{\lambda_{\max}(R_{u,k})}, \tag{20}$$

which is the same as that derived in [24].

Remark 4: We can conclude from (13) that even in the presence of fading channels, it is possible to ensure convergence of the ILMS algorithm in the mean, by choosing μ_k sufficiently small for all nodes k. Furthermore, the range that the step size can be chosen depends only on the mean of the channel fading statistics. This implies that if the channel gain can be estimated so that the resulting deviation gives $m_k = 1$, then the step size can be chosen to be the same as in the ideal link case. Furthermore, by comparing (13) with (20), we see that, in the presence of fading channels, the allowable range of step sizes for stability in mean can increase or decrease depending on the average channel gain.

Proposition 2: Suppose that Assumptions 1 and 2 hold, and that μ_k satisfies (13) for all $k = 1, \ldots, N$. Then, for the ILMS algorithm over fading channels, we have

$$\begin{aligned}
& \lim_{i\to\infty} \mathbb{E}[\widetilde{\boldsymbol{w}}_{k,i}] \\
& = (I - \mathcal{M})^{-1} \sum_{n=1}^{N}\left((1 - m_n)J_n \prod_{\ell=n+1}^{N} m_\ell J_\ell\right)w^o.
\end{aligned}$$

Proof: At steady state, we have $\lim_{i\to\infty} \mathbb{E}[\widetilde{\boldsymbol{w}}_{k,i}] = \lim_{i\to\infty} \mathbb{E}[\widetilde{\boldsymbol{w}}_{k,i-1}]$ for every node k. Taking the limit as $i \to \infty$ in (16), we obtain the desired result. ∎

Remark 5: From the proof of Proposition 2, it follows that if $m_k = 1$ for all nodes k, then the ILMS algorithm is asymptotically unbiased, which is consistent with the result for the mean convergence of the ILMS algorithm over ideal links or links with only additive noise [30], [31]. Furthermore, we see that if channel estimation achieves an error with average mean $m_k = 1$ for all channels, then the ILMS algorithm will also produce an asymptotically unbiased estimate. On the other hand, if $m_k \neq 1$ for some of the nodes k, then it is possible for the ILMS algorithm to produce an estimate that is asymptotically biased. From Proposition 2, this bias is the same for all nodes in the network. This is reminiscent of the convergence result for the ideal link case, which states that all the estimators $\boldsymbol{w}_{k,i}$ converge in the mean to w^o.

B. Steady State Mean-Square Performance Analysis

We now consider the steady state mean-square performance of the ILMS algorithm over fading channels. We are interested to quantify the performance using the following metrics at every node k:

$$\eta_k \triangleq \lim_{i \to \infty} \mathbb{E}\left[\|\widetilde{\boldsymbol{w}}_{k-1,i}\|^2\right] \quad \text{(MSD)} \tag{21}$$

$$\zeta_k \triangleq \lim_{i \to \infty} \mathbb{E}\left[\|\widetilde{\boldsymbol{w}}_{k-1,i}\|^2_{R_{u,k}}\right] \quad \text{(EMSE)} \tag{22}$$

$$\xi_k \triangleq \lim_{i \to \infty} \mathbb{E}\left[|\boldsymbol{e}_k(i)|^2\right] = \zeta_k + \sigma_{v,k}^2 \quad \text{(MSE)} \tag{23}$$

In general, to derive the above steady state performance metrics, we need to evaluate quantities of the form $\mathbb{E}\left[\|\widetilde{\boldsymbol{w}}_{k,i}\|^2_{\Sigma_k}\right]$ where Σ_k is a positive semi-definite Hermitian matrix. To this end, we consider the weight vector update equation given by (12). Let

$$C_{k,i} = m_k J_k C_{k-1,i} + (1-m_k) J_k, \tag{24}$$

where $C_{0,i} = C_{N,i-1}$, and $C_{0,1} = I$. Note that the matrix $C_{k,i}$ is such that

$$\mathbb{E}\left[\widetilde{\boldsymbol{w}}_{k,i}\right] = C_{k,i} w^o. \tag{25}$$

By equating the weighted norm of both sides of (12), taking expectations and using Assumptions 1 and 2, and (25), we obtain the following recursive relationship:

$$\mathbb{E}\left[\|\widetilde{\boldsymbol{w}}_{k,i}\|^2_{\Sigma_k}\right] = \mathbb{E}\left[\|\widetilde{\boldsymbol{w}}_{k-1,i}\|^2_{\Sigma'_k}\right] + \mu_k^2 \sigma_{v,k}^2 \mathbb{E}\left[\|\boldsymbol{u}_{k,i}\|^2_{\Sigma_k}\right]$$
$$+ \mathbb{E}\left[\|\boldsymbol{q}_{k,i}\|^2_{G_k}\right] + \|w^o\|^2_{T_k + H_{k,i}}, \tag{26}$$

where

$$G_k = \Sigma_k - \mu_k \mathbb{E}\left[\Sigma_k \boldsymbol{u}_{k,i}^* \boldsymbol{u}_{k,i} + \boldsymbol{u}_{k,i}^* \boldsymbol{u}_{k,i} \Sigma_k\right] \tag{27}$$
$$+ \mu_k^2 \mathbb{E}\left[\|\boldsymbol{u}_{k,i}\|^2_{\Sigma_k} \boldsymbol{u}_{k,i}^* \boldsymbol{u}_{k,i}\right]$$

$$\Sigma'_k = s_k G_k \tag{28}$$

$$T_k = (1 - 2m_k + s_k) G_k \tag{29}$$

$$H_{k,i} = (m_k - s_k)(C_{k-1,i} G_k + G_k C_{k-1,i}). \tag{30}$$

In order to compute all the moments that appear in the recursive equation (26) and to obtain closed-form expressions, we now make the following assumption regarding the regression vectors $\boldsymbol{u}_k$, for all nodes $k = 1, \ldots, N$.

Assumption 3: For each $k = 1, \ldots, N$, the distribution of $\boldsymbol{u}_k$ is a Gaussian distribution with

$$R_{u,k} = U_k \Lambda_k U_k^*, \tag{31}$$

where Λ_k is a diagonal matrix with diagonal elements being the eigenvalues of the correlation matrix $R_{u,k}$, and U_k is unitary matrix.

Making use of Assumption 3, we further define the following transformed quantities:

$$\begin{aligned}
\bar{\boldsymbol{w}}_{k,i} &= U_k^* \widetilde{\boldsymbol{w}}_{k,i}, & \bar{\Sigma}_k &= U_k^* \Sigma_k U_k, \\
\bar{\Sigma}'_k &= U_k^* \Sigma'_k U_k, & \bar{\boldsymbol{u}}_{k,i} &= \boldsymbol{u}_{k,i} U_k, \\
\bar{T}_k &= U_k^* T_k U_k, & \bar{H}_{k,i} &= U_k^* H_{k,i} U_k, \\
\bar{w}^o &= U_k^* w^o, & \bar{C}_{k-1,i} &= U_k^* C_{k-1,i} U_k, \\
\bar{Q}_k &= U_k^* Q_k U_k, & D &= \bar{w}^o \bar{w}^{o*} = w^o w^{o*}.
\end{aligned}$$

From the above definitions, equation (26) can now be rewritten in the following equivalent form

$$\mathbb{E}\left[\|\bar{\boldsymbol{w}}_{k,i}\|^2_{\bar{\Sigma}_k}\right] = \mathbb{E}\left[\|\bar{\boldsymbol{w}}_{k-1,i}\|^2_{\bar{\Sigma}'_k}\right] + \mu_k^2 \sigma_{v,k}^2 \mathbb{E}\left[\|\bar{\boldsymbol{u}}_{k,i}\|^2_{\bar{\Sigma}_k}\right]$$
$$+ \mathbb{E}\left[\|\bar{\boldsymbol{q}}_{k,i}\|^2_{\bar{G}_k}\right] + \|\bar{w}^o\|^2_{\bar{T}_k + \bar{H}_{k,i}} \tag{32}$$

where in (32) we have

$$\begin{aligned}
\bar{G}_k &= \bar{\Sigma}_k - \mu_k \mathbb{E}\left[\bar{\Sigma}_k \bar{\boldsymbol{u}}_{k,i}^* \bar{\boldsymbol{u}}_{k,i} + \bar{\boldsymbol{u}}_{k,i}^* \bar{\boldsymbol{u}}_{k,i} \bar{\Sigma}_k\right] \\
&\quad + \mu_k^2 \mathbb{E}\|\bar{\boldsymbol{u}}_{k,i}\|^2_{\bar{\Sigma}_k} \bar{\boldsymbol{u}}_{k,i}^* \bar{\boldsymbol{u}}_{k,i} \\
&= \bar{\Sigma}_k - \mu_k(\bar{\Sigma}_k \Lambda_k + \Lambda_k \bar{\Sigma}_k) \\
&\quad + \mu_k^2(\Lambda_k \text{Tr}[\bar{\Sigma}_k \Lambda_k] + \Lambda_k \bar{\Sigma}_k \Lambda_k) \tag{33}
\end{aligned}$$

$$\bar{\Sigma}'_k = s_k \bar{G}_k \tag{34}$$

$$\bar{T}_k = (1 - 2m_k + s_k) \bar{G}_k$$

$$\bar{H}_{k,i} = (m_k - s_k)(\bar{C}_{k-1,i} \bar{G}_k + \bar{G}_k \bar{C}_{k-1,i}) \tag{35}$$

Further algebraic manipulations of (32) yields

$$\mathbb{E}\left[\|\bar{\boldsymbol{w}}_{k,i}\|^2_{\bar{\Sigma}_k}\right] = \mathbb{E}\left[\|\bar{\boldsymbol{w}}_{k-1,i}\|^2_{\bar{\Sigma}'_k}\right] + \mu_k^2 \sigma_{v,k}^2 \text{Tr}[\Lambda_k \bar{\Sigma}_k]$$
$$+ \text{Tr}[\bar{Q}_k \bar{G}_k] + \text{Tr}[D\bar{T}_k] + \text{Tr}[D\bar{H}_{k,i}]. \tag{36}$$

To derive (21)-(23), we only need to consider the case where $\bar{\Sigma}_k$ is a diagonal matrix. In this case, matrix $\bar{\Sigma}'_k$ is also a diagonal matrix. We let

$$\bar{\sigma}_k \triangleq \text{diag}\{\bar{\Sigma}_k\}, \quad \bar{\sigma}'_k \triangleq \text{diag}\{\bar{\Sigma}'_k\}, \quad \lambda_k \triangleq \text{diag}\{\Lambda_k\}, \tag{37}$$

and

$$\bar{F}_k = I - \mu_k X_k + \mu_k^2 Y_k, \tag{38}$$

with $X_k = 2\Lambda_k$ and $Y_k = \Lambda_k^2 + \lambda_k \lambda_k^T$. The $M \times M$ matrix $\bar{F}_k$ contains the statistics of data local to node k. We then have

$$\mathbb{E}\left[\|\bar{\boldsymbol{w}}_{k,i}\|^2_{\bar{\sigma}_k}\right] = \mathbb{E}\left[\|\bar{\boldsymbol{w}}_{k-1,i}\|^2_{\bar{\sigma}'_k}\right] + g_{k,i} \bar{\sigma}_k, \tag{39}$$

where $g_{k,i}$ and $\bar{\sigma}'_k$ are given respectively by

$$\begin{aligned}
g_{k,i} &= \mu_k^2 \sigma_{v,k}^2 \lambda_k^T + \text{diag}\{\bar{Q}_k\}^T \bar{F}_k \\
&\quad + (1 - 2m_k + s_k)\text{diag}\{D\}^T \bar{F}_k \\
&\quad + 2(m_k - s_k)\text{diag}\{D\}^T \bar{C}_{k-1,i} \bar{F}_k, \\
\bar{\sigma}'_k &= s_k \bar{F}_k \bar{\sigma}_k. \tag{40}
\end{aligned}$$

We next use (39) to derive conditions that guarantee convergence in the mean-square sense for the ILMS algorithm with fading channels.

Proposition 3: Under assumptions 1, 2 and 3, the ILMS algorithm over fading channels converges in the mean-square sense if the step sizes μ_k are chosen to be sufficiently small so that the

$$s_k \rho(\bar{F}_k) < 1. \tag{41}$$

Proof: From (39) and using the results in Section 6.9 of [24], the ILMS algorithm with fading channels is stable in the mean-square sense if matrix $\bar{F}'_k \triangleq s_k \bar{F}_k$ is a stable matrix. Therefore, if we select the step size μ_k to be sufficiently small so that $\rho(\bar{F}'_k) < 1$, then stability in the mean-square sense is guaranteed. The proof is now complete. ∎

Remark 6: Suppose that $s_k = s$ for all nodes k, and the step sizes μ_k are fixed. Then, if s is sufficiently large, the left

hand side of (39) diverges and we no longer have mean-square stability. This shows that deteriorating fading conditions have detrimental impact on the ILMS algorithm, and care should be taken to adjust the step sizes according to Proposition 3.

Assuming that step sizes are chosen sufficiently small, and by letting $i \to \infty$, the recursive equation (39) at steady-state gives

$$\mathbb{E}\left[\|\bar{\boldsymbol{w}}_{k,\infty}\|^2_{\bar{\sigma}_k}\right] = \mathbb{E}\left[\|\bar{\boldsymbol{w}}_{k-1,\infty}\|^2_{\bar{\sigma}'_k}\right] + g_k\bar{\sigma}_k, \qquad (42)$$

where

$$\begin{aligned}
g_k &= \mu_k^2\sigma_{v,k}^2\lambda_k^T + \text{diag}\{\bar{Q}_k\}^T\bar{F}_k \\
&\quad + (1 - 2m_k + s_k)\text{diag}\{D\}^T\bar{F}_k \\
&\quad + 2(m_k - s_k)\text{diag}\{D\}^T\bar{C}_{k-1,\infty}\bar{F}_k, \qquad (43)
\end{aligned}$$

and using (14) and (25), $\bar{C}_{k-1,\infty}$ in (43) is given by

$$\begin{aligned}
\bar{C}_{k-1,\infty} = U_k^*\Bigg(&(I - \mathcal{M})^{-1}\times \\
&\sum_{n=1}^{N}\left((1 - m_n)J_n\prod_{\ell=n+1}^{N}m_\ell J_\ell\right)\Bigg)U_k. \quad (44)
\end{aligned}$$

We observe that (42) shows how $\mathbb{E}\left[\|\bar{\boldsymbol{w}}_{k,\infty}\|^2_{\bar{\sigma}_k}\right]$ evolves through the network, which in its current form makes it difficult to derive the desired metrics (21)-(23) directly. In fact, we have to find a recursive equation that reveals how $\mathbb{E}\left[\|\bar{\boldsymbol{w}}_{k,i}\|^2_{\bar{\sigma}_k}\right]$ evolves in time. By iterating (39), and using $\boldsymbol{w}_{0,i+1} = \boldsymbol{w}_{N,i}$, we can obtain a set of N coupled equations. With suitable manipulation of these equations, along with proper selections of $\bar{\sigma}_k$, it is possible to solve the resulting equalities to derive the desired metrics. Following the argument given in [12], we can derive the required metrics in a similar way. For completeness, we provide an outline of the proof in Appendix A.

Proposition 4: Under Assumptions 1, 2 and 3, the steady state performance of the ILMS algorithm over fading channels for each node k in the mean-square sense is given by the following expressions:

$$\eta_k = a_k(I - \Pi_{k,1})^{-1}\mathbb{1} \qquad (45)$$
$$\zeta_k = a_k(I - \Pi_{k,1})^{-1}\lambda_k \qquad (46)$$
$$\xi_k = \zeta_k + \sigma_{v,k}^2 \qquad (47)$$

where

$$\Pi_{k,l} \triangleq \left(\prod_{k=1}^{N}s_k\right)\left(\bar{F}_{k+l-1}\bar{F}_{k+l}\cdots\bar{F}_N\bar{F}_1\cdots\bar{F}_{k-1}\right), \quad (48)$$
$$a_k \triangleq g_k\Pi_{k,2} + g_{k+1}\Pi_{k,3} + \ldots + g_{k-2}\Pi_{k,N} + g_{k-1}, \quad (49)$$

where $l = 1, \cdots, N$ and all the subscripts are in $\mod N$.

Remark 7: Note that $g'_k = \mu_k^2\sigma_{v,k}^2\lambda_k^T$ is the equivalent expression for (43) for the original ILMS algorithm over ideal communication links (see equation (55) in [12]). By comparing (43) with equation (55) in [12], we can see that we can model the effect of fading channels as additional terms to g'_k with

$$\begin{aligned}
g_k = g'_k &+ \text{diag}\{\bar{Q}_k\}^T\bar{F}_k + (1 - 2m_k + s_k)\text{diag}\{D\}^T\bar{F}_k \\
&+ 2(m_k - s_k)\text{diag}\{D\}^T\bar{C}_{k-1,\infty}\bar{F}_k. \qquad (50)
\end{aligned}$$

The first additional term on the right hand side of (50) is due to the additive noise of the fading channel. The rest of the additional terms are due to statistics of the fading channel.

C. Dependence of MSD, EMSE and MSE on Channel Gain Variances

In this subsection, we investigate the dependence of the MSD, EMSE and MSE on the channel gain variance or second order moment by making some assumptions in order to simplify the analysis. We show that the MSD, EMSE and MSE are non-decreasing functions of s_k. To do this, we adopt the same Gaussian model used in the previous subsection, with the further assumptions that

1) $m_k = 1, \quad R_{u,k} = \lambda I, \quad Q_k = \sigma_{c,k}^2I$, and
2) $\mu_k = \mu$ and μ is sufficiently small so that (41) holds, and $\bar{F}_k$ can be approximated as

$$\bar{F}_k \approx I - \mu X = (1 - 2\mu\lambda)I.$$

Since $\bar{F}_k$ is now a diagonal matrix, the matrix $\Pi_{k,\ell} \approx \bar{F}_1\bar{F}_2\cdots\bar{F}_N$ is also diagonal, and can be approximated as

$$\Pi_{k,\ell} \approx s_p(1 - 2\mu\lambda)^N I, \qquad (51)$$

with $s_p \triangleq \prod_{k=1}^{N}s_k$. Using (51) we have

$$I - \Pi_{k,\ell} \approx \left(1 - s_p(1 - 2\mu\lambda)^N\right)I \qquad (52)$$

Similarly, we can also obtain approximations for g_k and a_k as follows:

$$\begin{aligned}
g_k &\approx (\mu^2\sigma_{v,k}^2\lambda)\mathbb{1}^T + (\sigma_{c,k}^2(1 - 2\mu\lambda))\mathbb{1}^T \\
&\quad + s_p(s_k - 1)(1 - 2\mu\lambda)\text{diag}\{D\}^T \\
&\approx \left(\mu^2\sigma_{v,k}^2\lambda + \sigma_{c,k}^2(1 - 2\mu\lambda)\right)\mathbb{1}^T \\
&\quad + s_p(s_k - 1)(1 - 2\mu\lambda)\text{diag}\{D\}^T \\
a_k &\approx s_p(1 - 2\mu\lambda)^N\left(\sum_{n=1}^{N}g_n - g_{k-1}\right) + g_{k-1} \\
&\approx \left(s_p(1 - 2\mu\lambda)^N\right)\sum_{n=1}^{N}g_n + (s_p(1 - 2\mu\lambda)^N)g_{k-1}. \quad (53)
\end{aligned}$$

Let $\hat{g}_k \triangleq g_k\mathbb{1}$. Since $\text{diag}\{D\} = [|w^o(1)|^2, \cdots, |w^o(M)|^2]^T$, we have

$$\begin{aligned}
\hat{g}_k = \mu^2\sigma_{v,k}^2\lambda &+ \sigma_{c,k}^2(1 - 2\mu\lambda) \\
&+ s_p(s_k - 1)(1 - 2\mu\lambda)\|w^o\|^2.
\end{aligned}$$

Replacing (52) and (53) in (45) we finally have

$$\eta_k \approx \left(\frac{1}{1 - s_p(1 - 2\mu\lambda)^N} - 1\right)\sum_{n=1}^{N}\hat{g}_n + \hat{g}_{k-1}. \qquad (54)$$

Similarly, we obtain

$$\zeta_k \approx \lambda\left(\frac{1}{1 - s_p(1 - 2\mu\lambda)^N} - 1\right)\sum_{n=1}^{N}\hat{g}_n + \lambda\hat{g}_{k-1}, \qquad (55)$$

and the approximation for the MSE ξ_k follows from (47). Since $\hat{g}_n$ is a non-decreasing and non-negative function of s_k for all k, from (54) and (55), we see that the MSD, EMSE and MSE are all non-decreasing functions of s_k (or equivalently
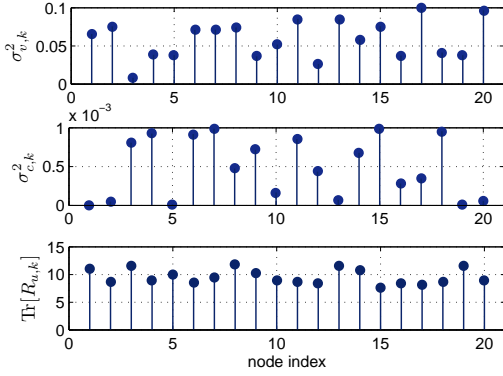
Fig. 2. Node profile and channel noise information: $\sigma_{v,k}^2$ (up), $\sigma_{c,k}^2$ (middle) and $\mathrm{Tr}(R_{u,k})$ (down).

the channel gain variances since we have assumed that the channel gain mean is 1).

Suppose further that $s_k = s \geq 1$ for all $k = 1, \ldots, N$, and the step size μ is chosen so that $s(1 - 2\mu\lambda)$ is a constant in $(0, 1)$ for all values of s. In this case, from (54) and (55), we obtain $\eta_k = \mathcal{O}(Ns^{N+1})$ and $\zeta_k = \mathcal{O}(Ns^{N+1})$.[1] This means that under the dB scale, the MSD, EMSE and MSE are approximately linear with respect to the channel gain second order moment. Simulations in Section IV are presented to verify our conclusions.

IV. SIMULATION RESULTS

To verify our theoretical performance analysis, we present some simulation results and compare them with the results in Section III. We assume a network composed of $N = 20$ nodes, where the nodes are connected via a ring topology as in the ILMS algorithm. The regressors $\boldsymbol{u}_{k,i}$ are generated as independent realizations of a Gaussian distribution with a covariance matrix $R_{u,k}$ whose eigenvalue spread is 4. The measurement data $\boldsymbol{d}_k(i)$ at each node k is generated by using the data model (1) where the parameter w^o is chosen to be $[1 \; 1 \; 1 \; 1]^T/2$, and the observation noise $\boldsymbol{v}_k(i)$ is drawn from a Gaussian distribution with variance $\sigma_{v,k}^2$ as shown in Figure 2. The additive channel noises are generated from Gaussian distributions with covariance matrix $Q_k = \sigma_{c,k}^2 I$, for $k = 1, \ldots, 20$. The values of $\sigma_{c,k}^2$ are shown in Figure 2. We generate the channel gains $\boldsymbol{h}_k(i)$ using a Rayleigh distribution with $m_k = \sqrt{2}/2$ for all values of k. To obtain the steady-state values of MSD, EMSE and MSE, we run the ILMS algorithm with 2000 iterations and average the last 200 samples. Finally, each steady-state value is obtained by averaging over 100 independent runs.

In Figure 3, we show the steady-state performance metrics MSD, EMSE and MSE as functions of the node index k when the step size $\mu = 0.02$. We can see that the simulated results closely match the theoretical results. In Figure 4, we have plotted the steady-state EMSE at node $k = 1$ versus step size μ. We see that in contrast to the ideal link case, the curve is
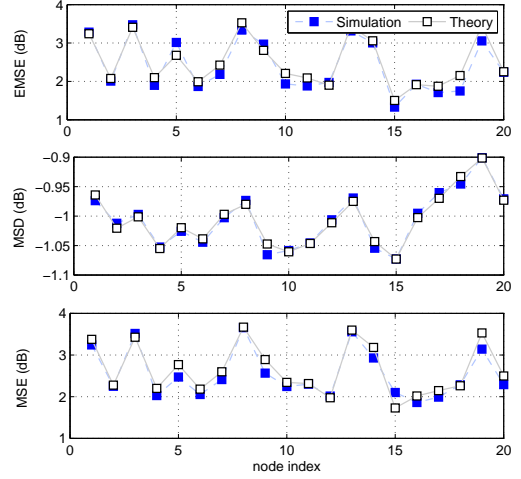
---

[1] A non-negative function $f(x)$ is $\mathcal{O}(g(x))$ if there exists $a > 0$ and x_0 such that for all $x \geq x_0$, $f(x) \leq ag(x)$.

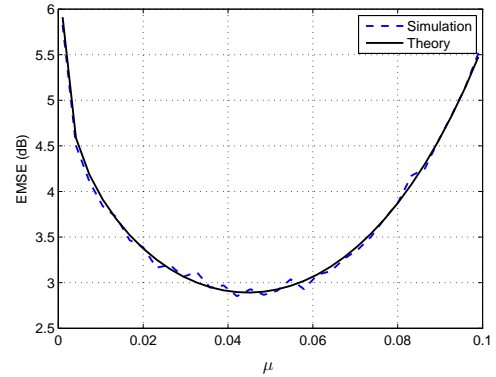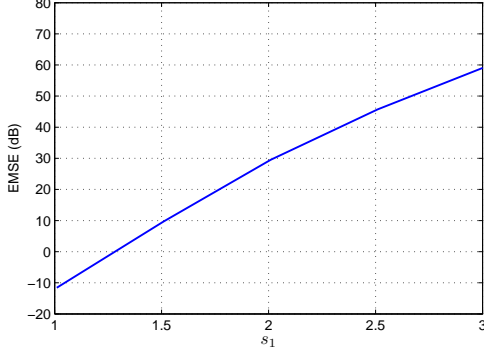Fig. 3. Steady-state curves versus for each individual node k, $\mu = 0.02$.

Fig. 4. Steady-state EMSE versus μ for node $k = 1$.

no longer a monotonically increasing function of the step-size μ.

To show how the channel statistics affect the estimation performance, we simulate the case where all channels have average gain 1. We further let $s_k = s$ for all $k = 1, \ldots, 20$, and evaluate the EMSE performance when s increases. We plot the steady-state EMSE at node $k = 1$ versus varying values of s in Figure 5. The step size is chosen as $\mu = 0.02$. It can be seen that as s increases, we get worse performance. On the other hand, as $s \to 0$, the performance of the ILMS algorithm over fading channels tend to the steady-state performance of the ILMS algorithm with additive noise links. Similar behavior is also seen for other steady-state metrics, i.e. MSD and MSE. This simulation also verifies our prediction of the linear relationship between EMSE and s in dB.

Although we assume that the regression vectors $\boldsymbol{u}_{k,i}$ are independent in this paper, our simulations indicate that the theoretical results also hold approximately for regression vectors with a shift structure [12]. To show this, we suppose that at each node k, and for all times i, the regression vector $\boldsymbol{u}_{k,i}$ can be expressed as

$$\boldsymbol{u}_{k,i} = [u_k(i), u_k(i-1), \cdots, u_k(i-M+1)],$$

Fig. 5. Steady-state EMSE versus channel gain second order moment s for node $k = 1$.

Fig. 7. Steady-state EMSE versus μ for node $k = 1$. We have considered regression vectors with shift structure.

where $u_k(i)$ is generated according to the following recursion

$$u_k(i) = \alpha_k u_k(i-1) + \beta_k \tau_k(i), \qquad (56)$$

which represents a first-order autoregressive (AR) process where with a pole at α_k. In (56), τ_k denotes a white, zero-mean, Gaussian random sequence with unit variance, while we choose $\alpha_k \in (0, 0.5]$ randomly and $\beta_k = \sqrt{\sigma_{u,k}^2(1 - \alpha_k^2)}$. In Figure 6, we show the steady-state performance metrics MSD, EMSE and MSE as functions of the node index k. We can see that for this case the simulated results have good match with the theoretical derivations. We also have plotted the steady-state EMSE at node $k = 1$ versus step size μ in Fig. 7.
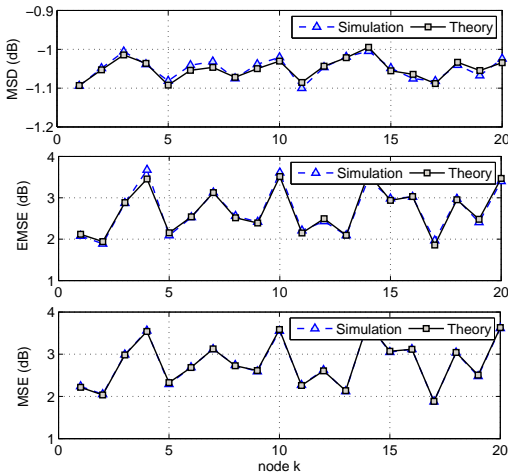

Fig. 6. Steady-state curves versus for each individual node k, $\mu = 0.02$.

V. CONCLUSION

In this paper, we have investigated the steady state performance of the ILMS algorithm when the links between nodes are fading channels, and we do not have perfect channel state information. Our analysis reveals some interesting behaviors, including (i) in the presence of fading channels, the ILMS algorithm is asymptotically biased, (ii) a dynamic range for the step-sizes for mean stability that depends only on the mean channel gain can be derived and (iii) mean-square stability
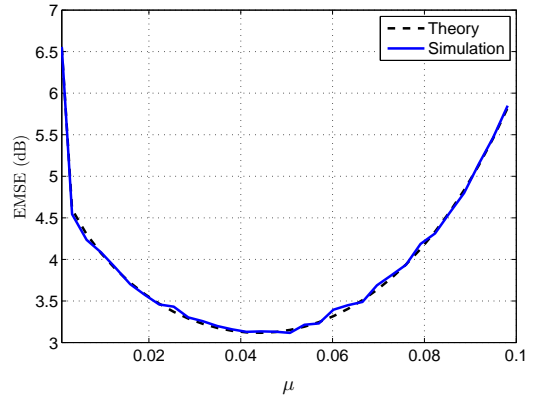
depends on the channel gain variances, with the MSD, EMSE and MSE being non-decreasing functions of these variances. We present simulation results to verify our theoretical analysis. Simulation results also show that the EMSE performance degrades as the fading channel gain variance increases.

In this paper, we focus on analyzing the effect of fading channels on incremental updates for the LMS filter. We believe that most of the insights obtained from this analysis are applicable to the more general class of diffusion LMS algorithms, whose performances under fading channels deserve further research.

APPENDIX A
OUTLINE PROOF OF PROPOSITION 4

In this appendix, we provide an outline proof of Proposition 4 for completeness. The proof steps are the same as those given in [12], except with additional m_k and s_k terms.

As we mentioned before, (42) is a coupled equation since it involves both $\mathbb{E}\left[\|\bar{\boldsymbol{w}}_{k,i}\|_{\bar{\sigma}_k}^2\right]$ and $\mathbb{E}\left[\|\bar{\boldsymbol{w}}_{k,i}\|_{\bar{\sigma}_{k-1}}^2\right]$, i.e., information from two spatial nodes. However, we need to find a recursive equation that reveals how $\mathbb{E}\left[\|\bar{\boldsymbol{w}}_{k,i}\|_{\bar{\sigma}_k}^2\right]$ evolves in time. We can exploit the ring (incremental) topology together with the weighting matrices to resolve this difficulty. To this end, let $\boldsymbol{p}_k \triangleq \bar{\boldsymbol{w}}_{k,\infty}$. By iterating (42) and using (40), we obtain the following coupled equalities

$$\mathbb{E}\left[\|\boldsymbol{p}_1\|_{\bar{\sigma}_1}^2\right] = \mathbb{E}\left[\|\boldsymbol{p}_N\|_{s_1\bar{F}_1\bar{\sigma}_1}^2\right] + g_1\bar{\sigma}_1,$$

$$\mathbb{E}\left[\|\boldsymbol{p}_2\|_{\bar{\sigma}_2}^2\right] = \mathbb{E}\left[\|\boldsymbol{p}_1\|_{s_2\bar{F}_2\bar{\sigma}_2}^2\right] + g_2\bar{\sigma}_2$$

$$\vdots$$

$$\mathbb{E}\left[\|\boldsymbol{p}_{k-2}\|_{\bar{\sigma}_{k-2}}^2\right] = \mathbb{E}\left[\|\boldsymbol{p}_{k-3}\|_{s_{k-2}\bar{F}_{k-2}\bar{\sigma}_{k-2}}^2\right] + g_{k-2}\bar{\sigma}_{k-2} \qquad (57)$$

$$\mathbb{E}\left[\|\boldsymbol{p}_{k-1}\|_{\bar{\sigma}_{k-1}}^2\right] = \mathbb{E}\left[\|\boldsymbol{p}_{k-2}\|_{s_{k-1}\bar{F}_{k-1}\bar{\sigma}_{k-1}}^2\right] + g_{k-1}\bar{\sigma}_{k-1} \qquad (58)$$

$$\vdots$$

$$\mathbb{E}\left[\|\boldsymbol{p}_N\|_{\bar{\sigma}_N}^2\right] = \mathbb{E}\left[\|\boldsymbol{p}_{N-1}\|_{s_N\bar{F}_N\bar{\sigma}_N}^2\right] + g_N\bar{\sigma}_N \qquad (59)$$

We can see from (58) that $\mathbb{E}\left[\|\boldsymbol{p}_{k-1}\|_{\bar{\sigma}_{k-1}}^2\right]$ in term of $\mathbb{E}\left[\|\boldsymbol{p}_{k-2}\|_{s_{k-1}\bar{F}_{k-1}\bar{\sigma}_{k-1}}^2\right]$. Now, if we choose the weighting vector $\bar{\sigma}_{k-2} = s_{k-1}\bar{F}_{k-1}\bar{\sigma}_{k-1}$ in (57) we get

$$\mathbb{E}\left[\|\boldsymbol{p}_{k-2}\|_{s_{k-1}\bar{F}_{k-1}\bar{\sigma}_{k-1}}^2\right] = \mathbb{E}\left[\|\boldsymbol{p}_{k-3}\|_{s_{k-1}s_{k-2}\bar{F}_{k-2}\bar{F}_{k-1}\bar{\sigma}_{k-1}}^2\right] \\ + s_{k-1}g_{k-2}\bar{F}_{k-1}\bar{\sigma}_{k-1} \quad (60)$$

Replacing (60) in (58) yields

$$\mathbb{E}\left[\|\boldsymbol{p}_{k-1}\|_{\bar{\sigma}_{k-1}}^2\right] = \mathbb{E}\left[\|\boldsymbol{p}_{k-3}\|_{s_{k-1}s_{k-2}\bar{F}_{k-2}\bar{F}_{k-1}\bar{\sigma}_{k-1}}^2\right] \\ + s_{k-1}g_{k-2}\bar{F}_{k-1}\bar{\sigma}_{k-1} \\ + g_{k-1}\bar{\sigma}_{k-1} \quad (61)$$

Iterating in this manner, we can obtain an expression for $\mathbb{E}\left[\|\boldsymbol{p}_{k-1}\|_{\bar{\sigma}_{k-1}}^2\right]$ as follows

$$E\left[\|\boldsymbol{p}_{k-1}\|_{\sigma_{k-1}}^2\right] = E\left[\boldsymbol{p}_{k-1}\|_{F_k'\ldots F_N' F_1'\ldots F_{k-1}'\sigma_{k-1}}^2\right] \\ + g_k F_{k+1}'\ldots F_N' F_1 \ldots F_{k-1}'\sigma_{k-1} \\ + g_{k+1} F_{k+2}'\ldots F_N' F_1'\ldots F_{k-1}'\sigma_{k-1} \\ + \cdots + g_{k-2}F_{k-1}'\sigma_{k-1} + g_{k-1}\sigma_{k-1} \quad (62)$$

Using (48) and (49) we can represent (62) in the following form

$$E\left[\|\boldsymbol{p}_{k-1}\|_{(I-\Pi_{k,1})\sigma_{k-1}}^2\right] = a_k\sigma_{k-1}. \quad (63)$$

Finally, we use equation (63) to derive the required metrics at node k. In fact, since we are free to select the weight vector σ_{k-1}, choosing $\sigma_{k-1} = (I - \Pi_{k,1})^{-1}\text{diag}\{I\}$ results in the expressions for the steady-state MSD as given by (45). Likewise, letting $\sigma_{k-1} = (I - \Pi_{k,1})^{-1}\lambda_k$ results in the expressions for the steady-state EMSE as given by (46).

REFERENCES

[1] I. F. Akyildiz, W. Su, Y. Sankarasubramaniam, and E. Cayirci, "A survey on sensor networks," *IEEE Commun. Mag.*, vol. 40, no. 8, pp. 102–114, Aug. 2002.

[2] I. Akyildiz and X. Wang, "A survey on wireless mesh networks," *IEEE Commun. Mag.*, vol. 43, no. 9, pp. S23–S30, 2005.

[3] N. Bulusu and S. Jha, *Wireless Sensor Networks: A Systems Perspective*. Artech House, 2005.

[4] C.-Y. Chong and S. P. Kumar, "Sensor networks: evolution, opportunities, and challenges," *Proc. IEEE*, vol. 91, no. 8, pp. 1247–1256, Aug. 2003.

[5] S.-Y. Tu and A. H. Sayed, "Adaptive decision making over complex networks," in *Proc. Asilomar Conference on Signals, Systems, and Computers*, Pacific Grove, CA, December 2012.

[6] W. P. Tay, J. N. Tsitsiklis, and M. Z. Win, "On the impact of node failures and unreliable communications in dense sensor networks," *IEEE Trans. Signal Processing*, vol. 56, no. 6, pp. 2535–2546, June 2008.

[7] J.-J. Xiao and Z.-Q. Luo, "Decentralized estimation in an inhomogeneous sensing environment," *IEEE Transactions on Information Theory*, vol. 51, no. 10, pp. 3564–3575, Oct. 2005.

[8] T. Aysal and K. Barner, "Constrained decentralized estimation over noisy channels for sensor networks," *Signal Processing, IEEE Transactions on*, vol. 56, no. 4, pp. 1398–1410, 2008.

[9] S. Boyd, A. Ghosh, B. Prabhakar, and D. Shah, "Randomized gossip algorithms," *Information Theory, IEEE Transactions on*, vol. 52, no. 6, pp. 2508–2530, 2006.

[10] I. Schizas, G. Mateos, and G. Giannakis, "Distributed LMS for consensus-based in-network adaptive processing," *Signal Processing, IEEE Transactions on*, vol. 57, no. 6, pp. 2365–2382, 2009.

[11] S. Kar and J. M. F. Moura, "Convergence rate analysis of distributed gossip (linear parameter) estimation: Fundamental limits and tradeoffs," *Selected Topics in Signal Processing, IEEE Journal of*, vol. 5, no. 4, pp. 674–690, 2011.

[12] C. G. Lopes and A. H. Sayed, "Incremental adaptive strategies over distributed networks," *IEEE Trans. Signal Processing*, vol. 55, no. 8, pp. 4064–4077, August 2007.

[13] A. H. Sayed and C. G. Lopes, "Distributed recursive least-squares strategies over adaptive networks," in *Proc. Asilomar Conf. Signals, Systems, Computers*, Monterey, CA, October 2006, pp. 233–237.

[14] L. Li, J. A. Chambers, C. G. Lopes, and A. H. Sayed, "Distributed estimation over an adaptive incremental network based on the affine projection algorithm," *IEEE Trans. Signal Process.*, vol. 58, no. 1, pp. 151–164, January 2010.

[15] N. Takahashi and I. Yamada, "Incremental adaptive filtering over distributed networks using parallel projection onto hyperslabs," *IEICE Technical Report*, vol. 108, pp. 17–22, 2008.

[16] C. G. Lopes and A. H. Sayed, "Randomized incremental protocols over adaptive networks," in *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing (ICASSP)*, Dallas, TX, March 2010, pp. 3514–3517.

[17] F. Cattivelli and A. H. Sayed, "Analysis of spatial and incremental LMS processing for distributed estimation," *IEEE Trans. on Signal Process.*, vol. 59, no. 4, pp. 1465–1480, April 2011.

[18] C. G. Lopes and A. H. Sayed, "Diffusion least-mean squares over adaptive networks: Formulation and performance analysis," *IEEE Trans. on Signal Process.*, vol. 56, no. 7, pp. 3122–3136, July 2008.

[19] F. S. Cattivelli, C. G. Lopes, and A. H. Sayed, "Diffusion recursive least-squares for distributed estimation over adaptive networks," *IEEE Trans. on Signal Process.*, vol. 56, no. 5, pp. 1865–1877, May 2008.

[20] F. S. Cattivelli and A. H. Sayed, "Multilevel diffusion adaptive networks," in *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing*, Taipei, Taiwan, April 2009.

[21] N. Takahashi, I. Yamada, and A. H. Sayed, "Diffusion least-mean-squares with adaptive combiners," in *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing*, Taipei, Taiwan, April 2009, pp. 2845–2848.

[22] S. Y. Tu and A. H. Sayed, "Mobile adaptive networks," *Selected Topics on Signal Processing*, vol. 5, no. 4, pp. 649–664, August 2011.

[23] J. Chen and A. H. Sayed, "Diffusion adaptation strategies for distributed optimization and learning over networks," *IEEE Trans. Signal Process.*, vol. 60, no. 8, pp. 4289–4305, August 2012.

[24] A. H. Sayed, "Diffusion adaptation over networks," in *E-Reference Signal Processing*, R. Chellapa and S. Theodoridis, Eds. Elsevier, 2013, to appear.

[25] A. H. Sayed, S.-Y. Tu, J. Chen, X. Zhao, and Z. Towfic, "Diffusion strategies for adaptation and learning over networks," *IEEE Signal Processing Magazine*, vol. 30, no. 3, pp. 155–171, May 2013.

[26] J. Tsitsiklis, D. Bertsekas, and M. Athans, "Distributed asynchronous deterministic and stochastic gradient optimization algorithms," *Automatic Control, IEEE Transactions on*, vol. 31, no. 9, pp. 803–812, 1986.

[27] A. Nedic and A. Ozdaglar, *Cooperative distributed multi-agent optimization*, Y. Eldar and D. Palomar, Eds. Cambridge, U. K.: Cambridge Univ. Press, 2009.

[28] S.-Y. Tu and A. Sayed, "Diffusion strategies outperform consensus strategies for distributed estimation over adaptive networks," *Signal Processing, IEEE Transactions on*, vol. 60, no. 12, pp. 6217–6234, 2012.

[29] C. G. Lopes and A. H. Sayed, "Distributed processing over adaptive networks," in *Proc. Adaptive Sensor Array Processing Workshop*. Lexington, MAMIT Lincoln Lab., Lexington, MA: MIT Lincoln Lab, June 2006.

[30] A. Khalili, M. A. Tinati, and A. Rastegarnia, "Performance analysis of distributed incremental LMS algorithm with noisy links," *Inter. Journal of Distributed Sensor Networks*, vol. 2011, pp. 1–10, 2011.

[31] ——, "Steady-state analysis of incremental LMS adaptive networks with noisy links," *IEEE Trans. Signal Processing*, vol. 59, no. 5, pp. 2416–2421, May 2011.

[32] A. Khalili, M. A. Tinati, A. Rastegarnia, and J. A. Chambers, "Transient analysis of diffusion least-mean squares adaptive networks with noisy channels," *Int. J. Adapt. Control Signal Process.*, vol. 26, no. 2, pp. 171–180, September 2011.

[33] ——, "Steady-state analysis of diffusion LMS adaptive networks with noisy links," *IEEE Trans. Signal Process.*, vol. 60, no. 2, pp. 974– 979, Feb. 2012.

[34] X. Zhao and A. H. Sayed, "Performance limits for distributed estimation over LMS adaptive networks," *IEEE Trans. Signal Process.*, vol. 60, no. 10, pp. 5107–5124, October 2012.

[35] X. Zhao, S.-Y. Tu, and A. H. Sayed, "Diffusion adaptation over networks under imperfect information exchange and non-stationary data," *IEEE Trans. on Signal Process.*, vol. 60, no. 7, pp. 3460–3475, July 2012.

[36] D. N. C. Tse and P. Viswanath, *Fundamentals of Wireless Communication*. Cambridge, UK: Cambridge University Press, 2005.

[37] R. Abdolee and B. Champagne, "Diffusion LMS algorithms for sensor networks over non-ideal inter-sensor wireless channels," in *Distributed Computing in Sensor Systems and Workshops (DCOSS), 2011 International Conference on*, 2011, pp. 1–6.

[38] R. Abdolee, B. Champagne, and A. H. Sayed, "Diffusion LMS strategies for parameter estimation over fading wireless channels," in *Proc. IEEE Int. Conf. on Communications*, Jun. 2013.

[39] A. H. Sayed, *Adaptive Filters*. NJ: Wiley, 2008.

[40] R. A. Horn and C. R. Johnson, Eds., *Matrix Analysis*. New York, NY, USA: Cambridge University Press, 1986.