PERFORMANCE LIMITS OF LMS-BASED ADAPTIVE NETWORKS

Xiaochuan Zhao and Ali H. Sayed

Department of Electrical Engineering, University of California, Los Angeles, CA 90095, Email: {xzhao, sayed}@ee.ucla.edu

ABSTRACT

In this work we analyze the mean-square performance of different strategies for adaptation over two-node least-mean-squares (LMS) networks. The results highlight some interesting properties for adaptive networks in comparison to centralized solutions. The analysis reveals that the adapt-then-combine (ATC) adaptive network algorithm can achieve lower excess-mean-square-error (EMSE) than a centralized solution that is based on either block or incremental LMS strategies with the same convergence rate.

***Index Terms*—** LMS, adaptive networks, diffusion, fusion center, incremental

1. INTRODUCTION

Consider an adaptive network consisting of only two nodes. We focus in this article on the case of two nodes because closed-form expressions for the mean-square performance of the nodes are possible in this case (under some assumptions on the data). These expressions facilitate the comparisons among the various algorithms. Despite this fact, it is worth noting that limiting the analysis to two nodes is still a challenging task. As is well-known in the adaptive filtering literature, studying the performance of a single stand-alone LMS filter is not trivial and generally requires certain assumptions on the data (this is because adaptive filters, by their very nature, are stochastic, nonlinear, and time-variant systems). When two adaptive nodes are allowed to interact with each other, as in the case in this paper, then the nodes end up influencing each other's behavior. For this reason, extending the performance analysis to the two nodes case is more demanding than the single node case.

We refer to the nodes as nodes 1 and 2. Both nodes are assumed to measure data that satisfy the linear regression model:

$$\boldsymbol{d}_k(i) = \boldsymbol{u}_{k,i} w^o + \boldsymbol{v}_k(i), \quad k = 1, 2 \tag{1}$$

where w^o is an $M \times 1$ unknown vector, $\boldsymbol{u}_{k,i}$ is the $1 \times M$ regression vector at time i, and $\boldsymbol{v}_k(i)$ is noise also at time i. All random variables are assumed to be zero-mean. The subscript k in (1) refers to the node number. The noise variance of node 2 is assumed to be less than that of node 1, i.e., $\sigma_{v,2}^2 < \sigma_{v,1}^2$. The nodes are interested in estimating the unknown parameter w^o. Assume initially that each node independently uses the least-mean squares (LMS) [1] algorithm to update its weight estimate, say, as (see Fig. 1):

$$\begin{cases} \text{node 1:} \ \boldsymbol{w}_{1,i} = \boldsymbol{w}_{1,i-1} + \mu \boldsymbol{u}_{1,i}^* (\boldsymbol{d}_1(i) - \boldsymbol{u}_{1,i} \boldsymbol{w}_{1,i-1}) \\ \text{node 2:} \ \boldsymbol{w}_{2,i} = \boldsymbol{w}_{2,i-1} + \mu \boldsymbol{u}_{2,i}^* (\boldsymbol{d}_2(i) - \boldsymbol{u}_{2,i} \boldsymbol{w}_{2,i-1}) \end{cases} \tag{2}$$

where μ denotes the step-size parameter. The performance of an adaptive algorithm is usually assessed in terms of its excess mean-square error (EMSE). If we introduce the a priori error

$$\boldsymbol{e}_{a,k}(i) \triangleq \boldsymbol{u}_{k,i}(w^o - \boldsymbol{w}_{k,i-1}) \tag{3}$$

This work was supported in part by NSF grants CCF-1011918, CCF-0942936, and ECS-0725441.

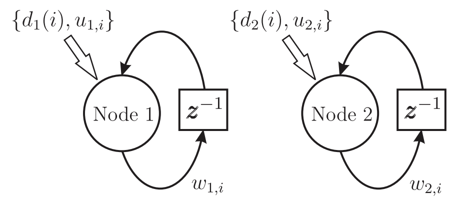

Fig. 1: Nodes 1 and 2 process the data independently by means of two local LMS filters.

then the EMSE is defined as [1]

$$\text{EMSE}_k \triangleq \lim_{i\to\infty} \mathsf{E}|\boldsymbol{e}_{a,k}(i)|^2 \tag{4}$$

It is known that for sufficiently small step-sizes, the EMSE of an LMS filter is given by [1]:

$$\text{EMSE}_k \approx \frac{1}{2}\mu\sigma_{v,k}^2 \mathsf{Tr}(R_{u,k}) \tag{5}$$

where $R_{u,k}$ is the covariance matrix of $\boldsymbol{u}_{k,i}$. Assuming a uniform regression data profile for both nodes, i.e., $R_{u,1} = R_{u,2} = R_u$, then expression (5) confirms the expected result that node 2 will achieve a better EMSE than node 1 because node 2 has lower noise variance than node 1. The interesting question that we wish to consider is whether it is possible to improve the EMSE performance for both nodes if they are allowed to cooperate with each other in some manner.

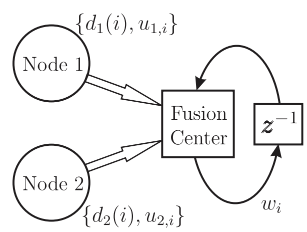

Fig. 2: Nodes 1 and 2 send their data to the fusion center for processing.

2. TWO CENTRALIZED ALGORITHMS

One standard way to realize cooperation is to connect the two nodes to a central fusion center that would collect their data and use the data to estimate w^o. The fusion center generally operates on the data in one of two ways. The first method is illustrated in Fig. 2 and we refer to it as block LMS. In this method, the weight estimate is updated as follows:

$$\boldsymbol{w}_i = \boldsymbol{w}_{i-1} + \mu' \begin{bmatrix} \boldsymbol{u}_{1,i} \\ \boldsymbol{u}_{2,i} \end{bmatrix}^* \left(\begin{bmatrix} \boldsymbol{d}_1(i) \\ \boldsymbol{d}_2(i) \end{bmatrix} - \begin{bmatrix} \boldsymbol{u}_{1,i} \\ \boldsymbol{u}_{2,i} \end{bmatrix} \boldsymbol{w}_{i-1} \right) \tag{6}$$

with step-size μ'. We see from (6) that at every iteration, the data

$\{\boldsymbol{d}_1(i), \boldsymbol{u}_{1,i}\}$ and $\{\boldsymbol{d}_2(i), \boldsymbol{u}_{2,i}\}$ from both nodes are blended together into vectors and used simultaneously by the fusion center to update $\boldsymbol{w}_{i-1}$ to $\boldsymbol{w}_i$. The second method that the fusion center can use to process the data is to apply two consecutive updates by using one data point at a time (see Fig. 3), i.e.,

$$\begin{cases} \boldsymbol{\phi}_{0,i} & \leftarrow & \boldsymbol{w}_{i-1} \\ \boldsymbol{\phi}_{1,i} & = & \boldsymbol{\phi}_{0,i} + \mu' \boldsymbol{u}_{1,i}^*(\boldsymbol{d}_1(i) - \boldsymbol{u}_{1,i}\boldsymbol{\phi}_{0,i}) \\ \boldsymbol{\phi}_{2,i} & = & \boldsymbol{\phi}_{1,i} + \mu' \boldsymbol{u}_{2,i}^*(\boldsymbol{d}_2(i) - \boldsymbol{u}_{2,i}\boldsymbol{\phi}_{1,i}) \\ \boldsymbol{w}_i & \leftarrow & \boldsymbol{\phi}_{2,i} \end{cases} \tag{7}$$

We see from (7) that the fusion center in this case first uses the data from node 1 to update $\boldsymbol{w}_{i-1}$ to an intermediate value $\boldsymbol{\phi}_{1,i}$, and then uses the data from node 2 to get $\boldsymbol{\phi}_{2,i}$, which is assigned to $\boldsymbol{w}_i$. Method (7) is actually a special case of the incremental LMS procedure proposed in [2].

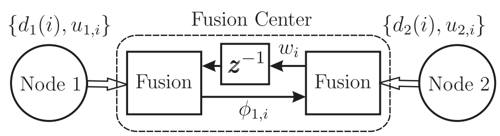

Fig. 3: Nodes 1 and 2 send their data to the fusion center where they are processed sequentially.

We observe from (6) and (7) that in going from $\boldsymbol{w}_{i-1}$ to $\boldsymbol{w}_i$, the block and incremental LMS algorithms employ two sets of data for each such update (while the conventional LMS recursions by the independent nodes in (2) employ one set of data for each update of their respective weight estimates). In order to ensure a fair comparison of the EMSE performance of the various algorithms, we shall set $\mu' = \mu/2$ so that the rates of convergence of all the algorithms considered in this paper are similar. Now, compared to the non-cooperative method (2) where the nodes act individually, the two fusion algorithms (6) and (7) can be shown to lead to improved EMSE performance (the arguments further ahead establish this conclusion among several other properties). The EMSE for block LMS (6) is defined as

$$\text{EMSE}_{\text{blk}} \triangleq \frac{1}{2} \lim_{i\to\infty} \mathsf{E}\|\boldsymbol{e}_{a,i}\|_2^2 \tag{8}$$

where the a priori error is now the two-dimensional vector:

$$\boldsymbol{e}_{a,i} \triangleq \begin{bmatrix} \boldsymbol{u}_{1,i} \\ \boldsymbol{u}_{2,i} \end{bmatrix} (w^o - \boldsymbol{w}_{i-1}) \tag{9}$$

Note that in (8) we are scaling by $1/2$ because the squared-Euclidean norm in (8) involves the sum of two error quantities. Likewise, the EMSE of the incremental implementation (7) is defined as:

$$\text{EMSE}_{\text{inc}} \triangleq \frac{1}{2} \lim_{i\to\infty} \left[\mathsf{E}|\boldsymbol{e}_{a,1}(i)|^2 + \mathsf{E}|\boldsymbol{e}_{a,2}(i)|^2\right] \tag{10}$$

where

$$\boldsymbol{e}_{a,k}(i) \triangleq \boldsymbol{u}_{k,i}(w^o - \boldsymbol{\phi}_{k-1,i}) \tag{11}$$

3. DIFFUSION ADAPTATION

We now consider fully decentralized algorithms of the diffusion type, which have been introduced in [3–6]. Our objective is to show that diffusion algorithms exploit the spatial diversity in the data more fully in a manner that can lead to better EMSE performance than the block and incremental algorithms when all algorithms converge at a similar rate.

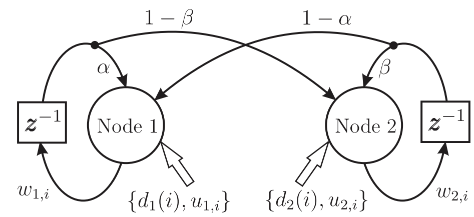

Fig. 4: A combine-then-adapt (CTA) diffusion adaptation step using combination coefficients $\{\alpha, 1-\alpha, \beta, 1-\beta\}$.

Diffusion algorithms consist of two steps: updating the weight estimate using local measurements (adaptation step) and aggregating the information from the neighbors (combination step). According to the order of these two steps, diffusion algorithms can be categorized into two classes: combine-then-adapt (CTA) (see Fig. 4):

$$\text{node 1:} \begin{cases} \boldsymbol{\phi}_{1,i-1} = \alpha \boldsymbol{w}_{1,i-1} + (1-\alpha)\boldsymbol{w}_{2,i-1} \\ \boldsymbol{w}_{1,i} = \boldsymbol{\phi}_{1,i-1} + \mu \boldsymbol{u}_{1,i}^*(\boldsymbol{d}_1(i) - \boldsymbol{u}_{1,i}\boldsymbol{\phi}_{1,i-1}) \end{cases} \tag{12}$$

$$\text{node 2:} \begin{cases} \boldsymbol{\phi}_{2,i-1} = \beta \boldsymbol{w}_{2,i-1} + (1-\beta)\boldsymbol{w}_{1,i-1} \\ \boldsymbol{w}_{2,i} = \boldsymbol{\phi}_{2,i-1} + \mu \boldsymbol{u}_{2,i}^*(\boldsymbol{d}_2(i) - \boldsymbol{u}_{2,i}\boldsymbol{\phi}_{2,i-1}) \end{cases} \tag{13}$$

and adapt-then-combine (ATC) (see Fig. 5):

$$\text{node 1:} \begin{cases} \boldsymbol{\phi}_{1,i} = \boldsymbol{w}_{1,i-1} + \mu \boldsymbol{u}_{1,i}^*(\boldsymbol{d}_1(i) - \boldsymbol{u}_{1,i}\boldsymbol{w}_{1,i-1}) \\ \boldsymbol{w}_{1,i} = \alpha \boldsymbol{\phi}_{1,i} + (1-\alpha)\boldsymbol{\phi}_{2,i} \end{cases} \tag{14}$$

$$\text{node 2:} \begin{cases} \boldsymbol{\phi}_{2,i} = \boldsymbol{w}_{2,i-1} + \mu \boldsymbol{u}_{2,i}^*(\boldsymbol{d}_2(i) - \boldsymbol{u}_{2,i}\boldsymbol{w}_{2,i-1}) \\ \boldsymbol{w}_{2,i} = \beta \boldsymbol{\phi}_{2,i} + (1-\beta)\boldsymbol{\phi}_{1,i} \end{cases} \tag{15}$$

where $\alpha, \beta \in [0,1]$ denote combination coefficients used by nodes 1 and 2. The ATC scheme (14)-(15) outperforms CTA (12)-(13) because it shares/diffuses updated information that is less noisy than CTA; see the analysis further ahead and also [4].

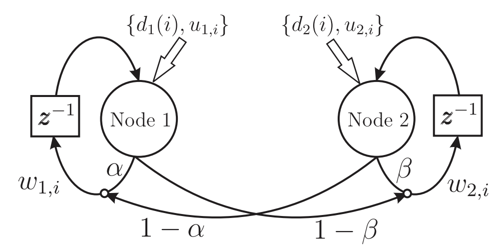

Fig. 5: An adapt-then-combine (ATC) diffusion adaptation step using combination coefficients $\{\alpha, 1-\alpha, \beta, 1-\beta\}$.

An important factor affecting the performance of diffusion LMS algorithms is the choice of the combination coefficients α and β. Different combination rules, such as uniform, Laplacian, maximum degree, Metropolis, relative degree, and relative degree-variance [4], have been proposed in the literature on graph models and networks. Apart from these static combination rules, where the coefficients are kept constant over time, adaptive rules are also possible. In the adaptive case, the combination weights can be adjusted regularly so that the network can respond to real-time node conditions [3] [7].

We now derive closed-form expressions for the mean-square performance of the LMS diffusion networks (12)-(15). The analysis highlights some interesting properties of adaptive networks in comparison to the (centralized) block and incremental algorithms (6) and (7).

4. PERFORMANCE ANALYSIS

We rely on the energy conservation framework of [1] to conduct the performance analysis. We introduce the following assumptions on the statistical properties of the data:

(a) The regression data $\boldsymbol{u}_{k,i}$ are temporally and spatially independent and identically distributed (i.i.d.) circular white Gaussian random variables with zero mean and diagonal covariance matrix $\sigma_u^2 I_M$.

(b) The noise signals $\boldsymbol{v}_k(i)$ are temporally and spatially i.i.d. circular white Gaussian random variables with zero mean and variances $\sigma_{v,k}^2$.

(c) The regression data $\boldsymbol{u}_{k,i}$ and noise signals $\boldsymbol{v}_k(j)$ are independent of each other for all i and j.

We start by recalling a classical result about the mean-square performance of a stand-alone LMS filter and use it to describe the EMSEs of the two independently-operating nodes in (2) when the regression data and noise signals are temporally i.i.d. circular white Gaussian random variables and independent of each other [1]:

$$\boxed{\mathrm{EMSE}_{\mathrm{ind},k} = \frac{M\mu^2\sigma_u^4\sigma_{v,k}^2}{1-\rho}, \quad k=1,2} \tag{16}$$

where

$$\rho \triangleq (1-\mu\sigma_u^2)^2 + M\mu^2\sigma_u^4 \tag{17}$$

The step-size μ is chosen to ensure $\rho < 1$ in order to guarantee mean square convergence. The resulting average EMSE of both nodes is

$$\boxed{\overline{\mathrm{EMSE}}_{\mathrm{ind}} = \frac{M\mu^2\sigma_u^4}{1-\rho}\cdot\frac{\sigma_{v,1}^2+\sigma_{v,2}^2}{2}} \tag{18}$$

4.1. EMSE of CTA LMS Networks

Studying the performance of diffusion networks is more challenging than stand-alone LMS filters due to the coupling between the filters as indicated in Fig. 4 and 5. Using energy conversation arguments, we can derive expressions for the EMSEs of both nodes in a CTA network. We omit the derivations for space limitations. When the step-size μ is sufficiently small so that $\sqrt{M}\mu\sigma_u^2 \ll 1$, we can show that the average EMSE of both nodes in a CTA network is given by:

$$\begin{aligned}\overline{\mathrm{EMSE}}_{\mathrm{cta}} &\triangleq \frac{1}{2}\left(\mathrm{EMSE}_{\mathrm{cta},1}+\mathrm{EMSE}_{\mathrm{cta},2}\right)\\ &\approx \frac{M\mu^2\sigma_u^4}{(2-\alpha-\beta)^2}\left[\frac{\sigma_{v,1}^2(1-\beta)^2+\sigma_{v,2}^2(1-\alpha)^2}{1-\rho}\right.\\ &\quad+\frac{(\sigma_{v,1}^2+\sigma_{v,2}^2)((1-\alpha)^2+(1-\beta)^2)}{2[1-\rho(\alpha+\beta-1)^2]}\\ &\quad-\frac{\sigma_{v,1}^2(1-\beta)^2+\sigma_{v,2}^2(1-\alpha)^2}{1-\rho(\alpha+\beta-1)}\\ &\quad\left.+\frac{(\sigma_{v,1}^2+\sigma_{v,2}^2)(1-\alpha)(1-\beta)}{1-\rho(\alpha+\beta-1)}\right]\end{aligned} \tag{19}$$

where $\mathrm{EMSE}_{\mathrm{cta},k}$ is the EMSE of node k. Without loss of generality, let us write $\sigma_{v,2}^2 = \gamma\sigma_{v,1}^2$ and assume $0<\gamma<1$. Then it can be verified that the network EMSE (19) is minimized when we choose

$$\alpha = \frac{\gamma}{1+\gamma} \quad \text{and} \quad \beta = \frac{1}{1+\gamma} \tag{20}$$

This choice coincides with the relative degree-variance rule proposed in [4]. The minimum value of (19) is then

$$\boxed{\overline{\mathrm{EMSE}}_{\mathrm{o-cta}} \approx M\mu^2\sigma_u^4\sigma_{v,1}^2\left(\frac{\rho}{1-\rho}\frac{\gamma}{1+\gamma}+\frac{1+\gamma}{2}\right)} \tag{21}$$

4.2. EMSE of ATC LMS Networks

Likewise, we can derive an expression for the EMSE of an ATC network:

$$\begin{aligned}\overline{\mathrm{EMSE}}_{\mathrm{atc}} &\triangleq \frac{1}{2}\left(\mathrm{EMSE}_{\mathrm{atc},1}+\mathrm{EMSE}_{\mathrm{atc},2}\right)\\ &\approx \frac{M\mu^2\sigma_u^4}{(2-\alpha-\beta)^2}\left[\frac{\sigma_{v,1}^2(1-\beta)^2+\sigma_{v,2}^2(1-\alpha)^2}{1-\rho}\right.\\ &\quad+\frac{(\sigma_{v,1}^2+\sigma_{v,2}^2)[(1-\alpha)^2+(1-\beta)^2](\alpha+\beta-1)^2}{2[1-\rho(\alpha+\beta-1)^2]}\\ &\quad-\frac{[(\sigma_{v,1}^2(1-\beta)^2+\sigma_{v,2}^2(1-\alpha)^2)](\alpha+\beta-1)}{1-\rho(\alpha+\beta-1)}\\ &\quad\left.+\frac{(\sigma_{v,1}^2+\sigma_{v,2}^2)(1-\alpha)(1-\beta)(\alpha+\beta-1)}{1-\rho(\alpha+\beta-1)}\right]\end{aligned} \tag{22}$$

where $\mathrm{EMSE}_{\mathrm{atc},k}$ is the EMSE of node k. We can again verify that the network EMSE (22) is minimized for the same choice (20). The resulting minimum EMSE is

$$\boxed{\overline{\mathrm{EMSE}}_{\mathrm{o-atc}} \approx \frac{M\mu^2\sigma_u^4\sigma_{v,1}^2}{1-\rho}\cdot\frac{\gamma}{1+\gamma}} \tag{23}$$

4.3. EMSEs of Block and Incremental LMS Networks

Using energy conservation arguments, we can also derive the EMSE of the block LMS network (6) (assuming $\mu' = \mu/2$) as:

$$\boxed{\mathrm{EMSE}_{\mathrm{blk}} = \frac{M\mu^2\sigma_u^4}{1-\rho'}\cdot\frac{\sigma_{v,1}^2+\sigma_{v,2}^2}{4}} \tag{24}$$

where

$$\rho' \triangleq (1-\mu\sigma_u^2)^2 + \frac{M}{2}\mu^2\sigma_u^4 \tag{25}$$

With regards to the incremental LMS network (7), we can use the results of [2] to get (also assuming $\mu' = \mu/2$):

$$\boxed{\mathrm{EMSE}_{\mathrm{inc}} \approx \frac{M\mu^2\sigma_u^4}{1-\rho'}\cdot\frac{\sigma_{v,1}^2+\sigma_{v,2}^2}{4}} \tag{26}$$

It is worth noting that, although (24) and (26) are identical for small μ, the incremental LMS algorithm actually outperforms block LMS because incremental LMS uses the intermediate estimate $\boldsymbol{\phi}_{1,i}$ during one step of the update in (7) while the block LMS in (6) does not [2] [8]. The intermediate estimate $\boldsymbol{\phi}_{1,i}$ is generally "less noisy" than $\boldsymbol{w}_{i-1}$ so that the incremental LMS gradually outputs a better estimate than the block LMS. However, we shall not distinguish between incremental LMS and block LMS in this work.

4.4. Performance Comparison

We summarize the mean-square performance of various strategies over two-node LMS networks in Table 1. We compare the performance of the algorithms based on the theoretical results in Table 2. The entries in Table 2 should be read from left to right. For example, the starred entry in the second row and fourth column is read as "O(ptimal)-ATC is better than incremental LMS".

Table 1: EMSEs of various strategies over two-node networks.

Type	Network EMSE	
O-ATC (14)-(15)	$c_1\sigma^2_{v,\text{harm}}$	(23)
O-CTA (12)-(13)	$c_1[\rho\sigma^2_{v,\text{harm}} + 2(1-\rho)\sigma^2_{v,\text{arth}}]$	(21)
Inc-LMS (6)	$c_2\sigma^2_{v,\text{arth}}$	(24)
Blk-LMS (7)	$c_2\sigma^2_{v,\text{arth}}$	(26)
Ind-LMS (2)	$2c_1\sigma^2_{v,\text{arth}}$	(18)

* $\sigma^2_{v,\text{harm}} \triangleq \frac{2\sigma^2_{v,1}\sigma^2_{v,2}}{\sigma^2_{v,1}+\sigma^2_{v,2}}$ and $\sigma^2_{v,\text{arth}} \triangleq \frac{\sigma^2_{v,1}+\sigma^2_{v,2}}{2}$.

† $c_1 \triangleq \frac{M\mu^2\sigma^4_u}{2(1-\rho)}$ and $c_2 \triangleq \frac{M\mu^2\sigma^4_u}{2(1-\rho')}$.

‡ ρ and ρ' are defined in (17) and (25), respectively.

§ $\sqrt{M}\mu\sigma^2_u \ll 1$ such that $c_1 \approx c_2$.

¶ The step-size for blk/inc LMS is $\mu' = \mu/2$.

Table 2: EMSE relationships for the various adaptation strategies.

	O-ATC	O-CTA	Inc-LMS	Blk-LMS	Ind-LMS
O-ATC	equal	slightly better	better*	better	better
O-CTA	slightly worse	equal	(28)	(28)	better
Inc-LMS	worse	(28)	equal	slightly better	better
Blk-LMS	worse	(28)	slightly worse	equal	better
Ind-LMS	worse	worse	worse	worse	equal

In addition, it can be verified that the performance of optimal ATC is better than any one of the individually-operating nodes:

$$\overline{\text{EMSE}}_{\text{o-atc}} < \text{EMSE}_{\text{ind},2} < \text{EMSE}_{\text{ind},1} \quad (27)$$

In this way, we find that optimal ATC is *optimal* among the other adaptive strategies listed in Table 2. However, as expressions (23) and (21) indicate, the performance of optimal ATC is only slightly better than that of optimal CTA for two-node networks, because ρ is usually close to one.

We can also verify that the relationship between the performance of optimal CTA and block/incremental LMS depends on ρ, $\sigma^2_{v,1}$, and $\sigma^2_{v,2}$:

$$\begin{cases} \overline{\text{EMSE}}_{\text{o-cta}} < \text{EMSE}_{\text{blk/inc}}, & \frac{4\sigma^2_{v,1}\sigma^2_{v,2}}{(\sigma^2_{v,1}+\sigma^2_{v,2})^2} < 2 - \frac{1}{\rho} \\ \overline{\text{EMSE}}_{\text{o-cta}} \geq \text{EMSE}_{\text{blk/inc}}, & \text{otherwise} \end{cases} \quad (28)$$

Similarly, the relationship between the performance of optimal CTA and individually-operating nodes depends on ρ, $\sigma^2_{v,1}$, and $\sigma^2_{v,2}$:

$$\begin{cases} \overline{\text{EMSE}}_{\text{o-cta}} < \text{EMSE}_{\text{ind},2}, & \sqrt{\frac{1-\rho}{1+\rho}} < \frac{\sigma^2_{v,2}}{\sigma^2_{v,1}} \\ \overline{\text{EMSE}}_{\text{o-cta}} \geq \text{EMSE}_{\text{ind},2}, & \text{otherwise} \end{cases} \quad (29)$$

5. SIMULATIONS RESULTS

We illustrate the theoretical results via simulations. The simulation profile is listed in Table 3. The weight vector w^o was selected randomly. Simulation results are shown in Fig. 6 from which we see that the theoretical results match well with simulation results. The optimal ATC (23) achieves the minimum network EMSE. The individual LMS (18) exhibits the worst EMSE level. The performance difference between optimal ATC and optimal CTA is negligible. And the performance difference between block LMS and incremental LMS is also negligible. Besides, it is worth noting that all algorithms converge at the same rate as a result of the normalization $\mu' = \mu/2$.

Table 3: Simulation profile.

M	μ	$\sigma^2_{v,1}$	$\sigma^2_{v,2}$	σ^2_u	# of iter.	# of trials
10	0.005	0.5	0.3	1	2000	1000

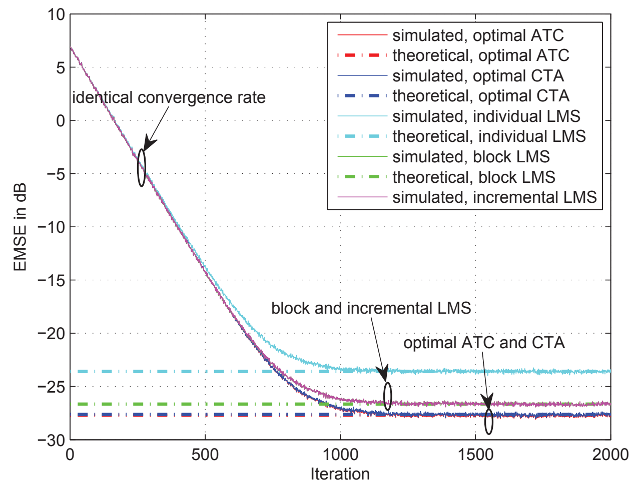

Fig. 6: EMSE comparison between different two-node LMS networks.

6. CONCLUSION

In this work we compared the EMSEs for various algorithms for adaptation over networks. We obtained expressions for the EMSEs of ATC and CTA networks. Based on the closed-form results, we were able to optimize the combination coefficients for ATC and CTA networks. The analysis shows that the relative degree-variance rule (20) is optimal for two-node LMS networks. Based on the analysis we further noted that an ATC network with optimal combining coefficients can achieve the lowest network EMSE when the step-sizes are carefully chosen for all algorithms such that they all converge at the same rate.

7. REFERENCES

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

[2] C. G. Lopes and A. H. Sayed, “Incremental adaptive strategies over distributed networks,” *IEEE Trans. Signal Process.*, vol. 48, pp. 223–229, Aug. 2007.

[3] ——, “Diffusion least-mean squares over adaptive networks: Formulation and performance analysis,” *IEEE Trans. Signal Process.*, vol. 56, pp. 3122–3136, Jul. 2008.

[4] F. Cattivelli and A. H. Sayed, “Diffusion LMS strategies for distributed estimation,” *IEEE Trans. Signal Process.*, vol. 58, pp. 1035–1048, Mar. 2010.

[5] S. Kar and J. M. F. Moura, “Distributed consensus algorithms in sensor networks: Link failures and channel noise,” *IEEE Trans. Signal Process.*, vol. 57, pp. 355–369, Jan. 2009.

[6] U. A. Khan, S. Kar, and J. M. F. Moura, “Distributed sensor localization in random environments using minimal number of anchor nodes,” *IEEE Trans. Signal Process.*, vol. 57, pp. 2000–2016, May 2009.

[7] N. Takahashi, I. Yamada, and A. H. Sayed, “Diffusion least-mean squares with adaptive combiners: Formulation and performance analysis,” *IEEE Trans. Signal Process.*, vol. 58, pp. 4795–4810, Sep. 2010.

[8] F. Cattivelli and A. H. Sayed, “Analysis of spatial and incremental LMS processing for distributed estimation,” *IEEE Trans. Signal Process.*, to appear, 2011.