This full text paper was peer reviewed at the direction of IEEE Communications Society subject matter experts for publication in the IEEE Globecom 2011 proceedings.

Adaptive Networks with Noisy Links

Sheng-Yuan Tu and Ali H. Sayed

Department of Electrical Engineering
University of California, Los Angeles, CA 90095
E-mail: {shinetu, sayed}@ee.ucla.edu

***Abstract*—In biological systems, animals exhibit organized behavior that arises from localized interactions. The interaction is implemented through information exchange, either directly or indirectly. Adaptive networks, consisting of a collection of nodes with learning abilities that interact with each other to solve distributed inference problems in real-time, are well-suited to model these kinds of behavior. Usually the information exchange between two nodes is imperfect and the data from neighbors are noisy. In this paper, we examine the effect of noisy communication links on network performance and derive an optimal strategy for adjusting the combination weights.**

***Index Terms*—Self-organization, adaptive networks, diffusion adaptation, noisy links.**

I. Introduction

Self-organization is observed in several physical and biological systems [1], [2]. For example, fish join together in schools and perform remarkable coherent motion to avoid attacks from predators and to improve foraging efficiency [3]; birds fly in V-formation while migrating [4]; and bees swarm towards a new hive [5]. In these cases, a global pattern of behavior emerges from localized interactions among the individual agents of the system. In earlier works [6]–[8], we used the framework of diffusion adaptation [9], [10] (see next section) to model these kinds of behavior.

In this paper, we examine the effect of noise over the communication links between the agents in an adaptive network. It was shown earlier in [11] that for perfect links, nodes benefit from cooperation. We argue further ahead that when the links are noisy, the nodes need to adjust their combination weights in accordance with the level of noise over the information links. If the weights are not adjusted, the noisy communication can degrade performance.

II. System Model

Consider a collection of N nodes distributed over a spatial domain. Two nodes in the network are said to be neighbors if they can share information. The set of neighbors of node k, including k itself, is called the neighborhood of k and is denoted by $\mathcal{N}_k$. The network of nodes would like to estimate some unknown column vector, w°, of size M. At every time instant, i, each node k is able to observe realizations of a scalar measurement, $d_k(i)$, and a $1 \times M$ row vector, $u_{k,i}$, with the latter arising from a stationary random process, $\boldsymbol{u}_{k,i}$, with zero mean and covariance matrix of the form $\sigma_u^2 I_M$. The measurements are assumed to be related to w° via a linear regression model of the form:

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

where $\boldsymbol{v}_k(i)$ is measurement noise with zero mean and variance $\sigma_{v,k}^2$. We assume that $\boldsymbol{v}_k(i)$ is spatially and temporally white and is independent of $\boldsymbol{u}_{l,i}$ for all l. In (1), we are using boldface letters to refer to random variables.

The objective of the network is to estimate w° by seeking the global minimizer of the cost function (2) below in a fully distributed manner and in real-time, where each node is allowed to interact only with its neighbors

$$J^{glob}(w) = \sum_{k=1}^{N} E|\boldsymbol{d}_k(i) - \boldsymbol{u}_{k,i} w|^2 \quad (2)$$

The letter E in (2) denotes the expectation operator. Several diffusion adaptation schemes for solving (3) in a distributed manner were proposed and studied in [9], [10]. One such scheme is the so-called Adapt-then-Combine (ATC) diffusion algorithm of [9]. It operates as follows:

$$\begin{cases} \psi_{k,i} = w_{k,i-1} + \mu u_{k,i}^* [d_k(i) - u_{k,i} w_{k,i-1}] & \text{(3a)} \\ \bar{\psi}_{k,i} = \psi_{k,i} + n_{k,i} & \text{(3b)} \\ w_{k,i} = a_{k,k}\psi_{k,i} + \sum_{l \in \mathcal{N}_k \setminus \{k\}} a_{k,l} \bar{\psi}_{l,i} & \text{(3c)} \end{cases}$$

In the first step (3a), node k uses its local measurements $\{d_k(i), u_{k,i}\}$ to update its existing estimate for w° from $w_{k,i-1}$ to some intermediate value, $\psi_{k,i}$, using an LMS-like update with step-size μ. In the second step (3b), node k exchanges its intermediate estimate with the other nodes in its neighborhood. Compared to [9], we are now assuming that the information exchange is contaminated by additive noise, $n_{k,i}$, in order to model imperfect channels. We assume that $n_{k,i}$ is a realization of a stationary random process, $\boldsymbol{n}_{k,i}$, with zero mean and covariance matrix $\sigma_{n,k}^2 I_M$. We also assume that $\boldsymbol{n}_{k,i}$ is temporally and spatially white and is independent of all other random variables. The final step (3c) is a combination step where node k combines the noisy information from its neighbors with its own estimate to obtain the updated estimate $w_{k,i}$ through a set of non-negative combination weights, $\{a_{k,l}\}$, satisfying

$$\sum_{l \in \mathcal{N}_k} a_{k,l} = 1 \text{ and } a_{k,l} = 0 \text{ if } l \notin \mathcal{N}_k \quad (4)$$

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

Other variants of the algorithm are possible, for example, by reversing the order of steps (3a) and (3c), we obtain the Combine-then-Adapt (CTA) diffusion solution as follows:

$$\begin{cases} \bar{w}_{k,i-1} = w_{k,i-1} + n_{k,i} & \text{(5a)} \\ \psi_{k,i-1} = a_{k,k} w_{k,i-1} + \sum_{l \in \mathcal{N}_k \setminus \{k\}} a_{k,l} \bar{w}_{l,i-1} & \text{(5b)} \\ w_{k,i} = \psi_{k,i-1} + \mu u_{k,i}^* [d_k(i) - u_{k,i} \psi_{k,i-1}] & \text{(5c)} \end{cases}$$

III. Mean-Square Performance

Define the error quantity $\tilde{\boldsymbol{w}}_{k,i} = w^\circ - \boldsymbol{w}_{k,i}$ and the corresponding network error vector $\tilde{\boldsymbol{w}}_i = \text{col}\{\tilde{\boldsymbol{w}}_{k,i}\}$. The network mean-square deviation (MSD) is defined as

$$\text{MSD} \triangleq \lim_{i \to \infty} \frac{1}{N} E \|\tilde{\boldsymbol{w}}_i\|^2 \tag{6}$$

where the notation $\|\cdot\|^2$ denotes the squared Euclidean norm of its argument. To examine the mean-square performance of the algorithm, we rely on the energy conservation approach of [12], [13]. Thus, let Σ denote an $NM \times NM$ positive-definite Hermitian matrix that we are free to choose. We partition Σ into $N \times N$ blocks:

$$\Sigma = [\Sigma_{kl}], \quad k, l = 1, 2, \ldots, N$$

where each submatrix Σ_{kl} is chosen as a diagonal matrix. We introduce the matrix operation "vdiag", which vectorizes the diagonal submatrices Σ_{kl}, as follows:

$$\begin{aligned} \sigma &= \text{vdiag}(\Sigma, N) \\ &\triangleq \text{col}\{\text{diag}(\Sigma_{11}), \ldots, \text{diag}(\Sigma_{1N}), \ldots, \\ &\qquad \text{diag}(\Sigma_{N1}), \ldots, \text{diag}(\Sigma_{NN})\} \end{aligned} \tag{7}$$

$$\Sigma = \text{vdiag}(\sigma, N) \tag{8}$$

In (8), we are also using "vdiag" to recover the original matrix Σ from its vectorized version σ. We further let either notation $\|x\|_\Sigma^2 = \|x\|_{\text{vdiag}(\sigma,N)}^2$ denote the squared weighted norm of a vector x. In the sequel, for compactness of presentation, we shall drop the vdiag$\{\cdot, N\}$ notation from the subscript and keep the vector σ only and thus write $\|x\|_\sigma^2$ to refer to the same squared weighted norm.

A. Variance Relation for ATC

We focus on the case of two nodes ($N = 2$) to highlight an interesting pattern of behavior; in subsequent work, we will extend the results to larger values of N. For $N = 2$, we denote the combination weights by $a_{11} = 1 - a$, $a_{12} = a$, $a_{22} = 1 - b$, and $a_{21} = b$ (see Fig. 1). In addition, we let $\sigma_{v,1}^2 = r\sigma_v^2$ and $\sigma_{v,2}^2 = \sigma_v^2$ for some positive scaling factor r. Similarly, let $\sigma_{n,1}^2 = s\sigma_n^2$ and $\sigma_{n,2}^2 = \sigma_n^2$. Moreover, for notational simplicity, we introduce the two variables:

$$\rho = 1 - 2\mu\sigma_u^2, \quad t = \sigma_n^2 / \sigma_v^2 \tag{9}$$

Note that t is the ratio between the variances of the two noise sources — one is measurement noise at the node (σ_v^2) and the other is noise over the link (σ_n^2).

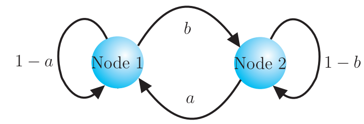

Fig. 1. Combination weights for a two-node network.

Starting from (3) and using model (1), we obtain the following recursion for $\tilde{\boldsymbol{w}}_i$:

$$\tilde{\boldsymbol{w}}_i = A(I - \mu \boldsymbol{U}_i)\tilde{\boldsymbol{w}}_{i-1} - \mu A \boldsymbol{g}_i - B \boldsymbol{n}_i \tag{10}$$

where

$$\begin{aligned} \boldsymbol{U}_i &= \text{diag}\{\boldsymbol{u}_{1,i}^* \boldsymbol{u}_{1,i}, \boldsymbol{u}_{2,i}^* \boldsymbol{u}_{2,i}\} \\ \boldsymbol{g}_i &= \text{col}\{\boldsymbol{u}_{1,i}^* \boldsymbol{v}_{1,i}, \boldsymbol{u}_{2,i}^* \boldsymbol{v}_{2,i}\} \\ \boldsymbol{n}_i &= \text{col}\{\boldsymbol{n}_{1,i}, \boldsymbol{n}_{2,i}\} \end{aligned}$$

and

$$A = \begin{bmatrix} 1-a & a \\ b & 1-b \end{bmatrix} \otimes I_M, \quad B = \begin{bmatrix} 0 & a \\ b & 0 \end{bmatrix} \otimes I_M$$

with $\otimes$ representing the Kronecker product operation. Following the energy conservation approach of [12], we can then establish the following weighted variance relation:

$$\begin{aligned} E\|\tilde{\boldsymbol{w}}_i\|_\Sigma^2 = {} & E\left(\|\tilde{\boldsymbol{w}}_{i-1}\|^2_{(I-\mu\boldsymbol{U}_i)A^T\Sigma A(I-\mu\boldsymbol{U}_i)}\right) \\ & + \mu^2 E[\boldsymbol{g}_i^* A^T \Sigma A \boldsymbol{g}_i] + E[\boldsymbol{n}_i^* B^T \Sigma B \boldsymbol{n}_i] \end{aligned} \tag{11}$$

Suppose the regressors $\{\boldsymbol{u}_{k,i}\}$ are independent over time and space. This assumption implies that $\boldsymbol{u}_{k,i}$ is independent of $\tilde{\boldsymbol{w}}_{i-1}$. For small step sizes, we appeal to the approximation:

$$E[(I - \mu\boldsymbol{U}_i)A^T \Sigma A (I - \mu \boldsymbol{U}_i)] \approx \rho A^T \Sigma A \tag{12}$$

In addition, we have

$$E[\boldsymbol{g}_i^* A^T \Sigma A \boldsymbol{g}_i] = \sigma_u^2 \sigma_v^2 h_{v,ATC}^T \sigma \tag{13}$$

$$E[\boldsymbol{n}_i^* B^T \Sigma B \boldsymbol{n}_i] = \sigma_n^2 h_{n,ATC}^T \sigma \tag{14}$$

where

$$h_{v,\text{ATC}} = \begin{bmatrix} (1-a)^2 r + a^2 \\ (1-a)br + a(1-b) \\ (1-a)br + a(1-b) \\ b^2 r + (1-b)^2 \end{bmatrix} \otimes \mathbf{1}, \;\; h_{n,\text{ATC}} = \begin{bmatrix} a^2 \\ 0 \\ 0 \\ b^2 s \end{bmatrix} \otimes \mathbf{1}$$

and $\mathbf{1}$ is a vector of size M with all entries equal to 1. From (12), (13), and (14), we can rewrite (11) as

$$E\|\tilde{\boldsymbol{w}}_i\|_\sigma^2 = E\|\tilde{\boldsymbol{w}}_{i-1}\|_{\rho\mathcal{A}\sigma}^2 + h_{\text{ATC}}^T \sigma \tag{15}$$

where

$$\mathcal{A} = A^T \otimes A^T \tag{16}$$

$$h_{\text{ATC}} = \sigma_v^2 \left[\frac{1}{2}\mu(1-\rho)h_{v,\text{ATC}} + t \cdot h_{n,\text{ATC}}\right] \tag{17}$$

and where we used the fact that vdiag$(A^T \Sigma A) = \mathcal{A}\sigma$.

B. Variance Relation for CTA

Similarly, starting from (5) and using model (1), we obtain the following weighted variance relation:

$$E\|\tilde{\boldsymbol{w}}_i\|^2_\Sigma = E\left(\|\tilde{\boldsymbol{w}}_{i-1}\|^2_{A^T(I-\mu\boldsymbol{U}_i)\Sigma(I-\mu\boldsymbol{U}_i)A}\right) + \mu^2 E[\boldsymbol{g}_i^*\Sigma\boldsymbol{g}_i] + E[\boldsymbol{n}_i^* B^T(I-\mu\boldsymbol{U}_i)\Sigma(I-\mu\boldsymbol{U}_i)B\boldsymbol{n}_i] \quad (18)$$

For small step sizes, we can rewrite (18) as

$$E\|\tilde{\boldsymbol{w}}_i\|^2_\sigma = E\|\tilde{\boldsymbol{w}}_{i-1}\|^2_{\rho\mathcal{A}\sigma} + h^T_{\text{CTA}}\sigma \quad (19)$$

where

$$h_{\text{CTA}} = \sigma_v^2\left[\frac{1}{2}\mu(1-\rho)h_{v,\text{CTA}} + t\cdot h_{n,\text{CTA}}\right] \quad (20)$$

with

$$\begin{aligned} h_{n,\text{CTA}} &= \rho\cdot h_{n,\text{ATC}} \\ h^T_{v,\text{CTA}} &= \begin{bmatrix} r & 0 & 0 & 1\end{bmatrix}\otimes\mathbf{1}^T \end{aligned} \quad (21)$$

C. Mean-Square Deviation

Since the ATC and CTA algorithms have similar variance relations in (15) and (19), we derive an expression for the network MSD for a general form of the variance relation as below and then specialize the results to ATC and CTA:

$$E\|\tilde{\boldsymbol{w}}_i\|^2_\sigma = E\|\tilde{\boldsymbol{w}}_{i-1}\|^2_{\mathcal{F}\sigma} + h^T\sigma \quad (22)$$

where $\mathcal{F}$ and h are of the form:

$$\mathcal{F} = \rho\mathcal{A} \quad (23)$$

$$h^T = \begin{bmatrix} h_1 & h_2 & h_3 & h_4\end{bmatrix}\otimes\mathbf{1}^T \quad (24)$$

Suppose the step size μ is chosen such that the eigenvalues of the matrix $\mathcal{F}$ are within the unit disc. Since A is a stochastic matrix, the spectral radius of $\mathcal{A}$ is one and we can choose μ such that $|\rho| < 1$, i.e., $0 < \mu\sigma_u^2 < 1$. Taking the limit of (22) as $i\to\infty$, the network MSD is obtained by choosing

$$\begin{aligned} \sigma &= (I-\mathcal{F})^{-1}\text{vdiag}(I_{2M}) \\ &= (I-\mathcal{F})^{-1}[(e_1+e_4)\otimes\mathbf{1}] \end{aligned} \quad (25)$$

so that

$$\text{MSD} = \frac{1}{2}h^T(I-\mathcal{F})^{-1}[(e_1+e_4)\otimes\mathbf{1}] \quad (26)$$

where e_i is a vector of size 4 with all entries equal to 0 except the ith entry, which is equal to 1. Now note that the eigenvalue decomposition for the matrix A^T is $A^T = V\Lambda V^{-1}$ where

$$V = \begin{bmatrix} b & 1 \\ a & -1\end{bmatrix}\otimes I_M, \quad \Lambda = \begin{bmatrix} 1 & 0 \\ 0 & 1-a-b\end{bmatrix}\otimes I_M$$

Therefore, the matrix $\mathcal{F}$ can be expressed as

$$\mathcal{F} = \rho(V\otimes V)(\Lambda\otimes\Lambda)(V^{-1}\otimes V^{-1}) \quad (27)$$

Some algebra shows that (26) becomes

$$\text{MSD} = \frac{M}{2(a+b)^2}\Big\{2\left[b^2h_1 + ab(h_2+h_3) + a^2h_4\right]\nu_1 + (a-b)\left[2bh_1 + (a-b)(h_2+h_3) - 2ah_4\right]\nu_2 + (a^2+b^2)\left[h_1 - (h_2+h_3) + h_4\right]\nu_3\Big\} \quad (28)$$

where

$$\nu_1 = \frac{1}{1-\rho}, \quad \nu_2 = \frac{1}{1-\rho\lambda}, \quad \nu_3 = \frac{1}{1-\rho\lambda^2}$$

and $\lambda = 1-a-b$. Note that, for $a = b$, we can simplify the MSD expression to

$$\text{MSD} = \frac{M}{4}\big[(h_1+h_2+h_3+h_4)\nu_1 + (h_1-h_2-h_3+h_4)\nu_3\big] \quad (29)$$

Therefore, using (17) and (20), we deduce the network MSD's for the ATC and CTA algorithms by substituting the value of the corresponding vectors h.

IV. Optimal Combination Weights

In this section, we determine a set of weights $\{a, b\}$ such that the network MSD given by (28) is minimized. To do so, we set the partial derivative of the MSD with respect to a and b to zero, i.e.,:

$$\frac{\partial\text{MSD}}{\partial a} = \frac{\partial\text{MSD}}{\partial b} = 0 \quad (30)$$

From the first equality in (30), we obtain that, for ATC and CTA algorithms, the optimal values for $\{a, b\}$ are the same and they should satisfy the relation:

$$\boxed{\begin{aligned} &\mu(1-\rho)\left[2(a-rb)(\nu_1-\nu_2) + (r+1)(a-b)(\nu_3-\nu_2)\right] \\ &+ 2\rho t[2(s+1)ab(b-a)(\nu_1-\nu_2) + (a+b)^2(a-sb)\nu_2] \\ &+ 2\rho t\left[(s+1)ab(b-a) + 2(a^3-sb^3)\right](\nu_3-\nu_2) = 0 \end{aligned}} \quad (31)$$

To examine the influence of noise due to information exchange, we consider the example corresponding to the case $r = s = 1$ (i.e., equal noise levels at both nodes for both measurement noise and information exchange noise). In this case, relation (31) between $\{a, b\}$ simplifies to

$$\boxed{a = b} \quad (32)$$

This result is expected since both nodes have the same noise levels. Setting the derivative of the MSD (29) with respect to a to zero, we find that the optimal value for a should satisfy the following fifth-order equation:

$$\boxed{\begin{aligned} &16\rho^2ta^5 - 32\rho^2ta^4 + 8\rho(3\rho-1)ta^3 + 10\rho(1-\rho)ta^2 \\ &+ 2(1-\rho)^2(t+\mu)a - \mu(1-\rho)^2 = 0 \end{aligned}} \quad (33)$$

Note that relation (33) holds for the ATC and CTA algorithm. Let the function on the left hand side of (33) be denoted by $f(a)$. In the following, we are going to show that there is a single value of a that minimizes the MSD and that this value for a lies between $(0, 0.5]$. We first establish some useful properties.

Lemma 1: Assume $r = s = 1$ and $a = b$. For $0 < \rho < 1$, $t \geq 0$ and $a \in [0, 1]$, the network MSD (29) and the function $f(a)$ satisfy the following properties:

1) $\frac{\partial\text{MSD}}{\partial a}\big|_{a=0} < 0$ and $\frac{\partial\text{MSD}}{\partial a}\big|_{a=1} > 0$.
2) $f(0) < 0$ and $f(0.5) \geq 0$ with equality to zero if, and only if, $t = 0$.

This full text paper was peer reviewed at the direction of IEEE Communications Society subject matter experts for publication in the IEEE Globecom 2011 proceedings.

3) $f(a) > 0$ if $0.5 < a \leq 1$.
4) $f'(a) > 0$ if $0 \leq a \leq 0.5$.

Proof:

1) We only examine the MSD for the CTA algorithm. A similar argument applies to the ATC algorithm. Taking the derivative of (29), with $h = h_{\text{CTA}}$, with respect to a, we obtain

$$\begin{aligned}\frac{\partial \text{MSD}_{\text{CTA}}}{\partial a} =& M\sigma_v^2\rho\Bigg[ta\left(\frac{1}{1-\rho}+\frac{1}{1-\rho\lambda^2}\right) \\ &- [\mu(1-\rho)+2t\rho a^2]\frac{1-2a}{(1-\rho\lambda^2)^2}\Bigg]\end{aligned} \quad (34)$$

Then,

$$\left.\frac{\partial \text{MSD}_{\text{CTA}}}{\partial a}\right|_{a=0} = M\sigma_v^2\rho\frac{-\mu}{1-\rho} < 0$$

$$\left.\frac{\partial \text{MSD}_{\text{CTA}}}{\partial a}\right|_{a=1} = M\sigma_v^2\rho[\mu(1-\rho)+2t]\frac{1}{(1-\rho)^2} > 0$$

2) From the expression (33) for f(a) we get:

$$f(0) = -\mu(1-\rho)^2 < 0$$

$$f(0.5) = \frac{1}{2}(2-\rho)t \geq 0$$

Note that since $0 < \rho < 1$, we would obtain $f(0.5) = 0$ if, and only if, $t = 0$.

3) Some algebra shows that the function $f(a)$ can be rewritten as

$$\begin{aligned}f(a) =& 2\rho(1-\rho)ta(1-a)(2a-1)\left[(2a-1)^2+2a\right] \\ &+ 16\rho ta^3(1-a)^2 + 2(1-\rho)ta \\ &+ \mu(1-\rho)^2(2a-1)\end{aligned} \quad (35)$$

Therefore, for $0.5 < a \leq 1$, $f(a) > 0$.

4) Taking the derivative of $f(a)$ and rearranging the result, we obtain

$$\begin{aligned}f'(a) =& 4\rho(1-\rho)ta\left[(1-2a)^2(3-5a)+2-a\right] \\ &+ 16\rho ta^2(1-a)(3-5a) + 2(1-\rho)^2(t+\mu)\end{aligned} \quad (36)$$

It is clear that $f'(a) > 0$ if $0 \leq a \leq 0.5$.

■

With these properties, we arrive at the following result.

Theorem 1: Assume $r = s = 1$ and $a = b$. There is a unique $a \in (0, 0.5]$ that minimizes the network MSD (29).

Proof: Since the MSD is a bounded function of a when $a \in [0, 1]$, there exists a value of a that minimizes the MSD. From the first property in the previous lemma, we see that the optimal a cannot be at the end points, 0 or 1, and has to satisfy equation (33). From the second and third properties, we conclude that there exists at least one value of $a \in (0, 0.5]$ satisfying (33) and for any value of a greater than 0.5, equation (33) does not hold. To establish the uniqueness of the minimizing a, from the fourth property, we know that $f(a)$ is a strictly increasing function of a when $0 \leq a \leq 0.5$. Therefore, there is only one value of $a \in (0, 0.5]$ satisfying (33) and that minimizes the MSD.

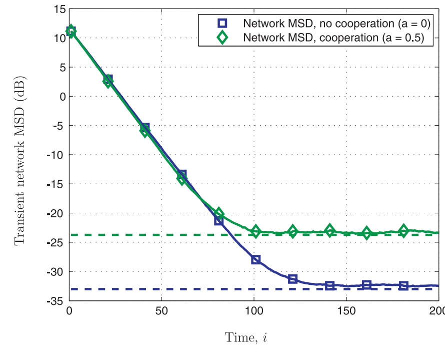

Fig. 2. Transient network MSD over time. The dashed lines indicate theoretical steady-state MSD.

■

From the theorem, we observe that, under the condition where both nodes have the same noise levels, the average strategy (i.e., $a = b = 1/2$) is the optimal strategy if, and only if, there is no information sharing error. If there is noise from information exchange, the nodes have to lower the value of the combination weight, a, i.e., a node has to place more weight on its own estimate.

V. Simulation Results

In this section, we present simulation results. The variance of the regressor vector is set to $\sigma_u^2 = 1$ and the variance of the measurement noise is set to $\sigma_v^2 = 0.01$. In Fig. 2, we show the transient network MSD for the CTA algorithm over time. The step size is $\mu = 0.05$ and the variance of noise due to information exchange is $\sigma_n^2 = 1.6 \times 10^{-3}$, i.e., $t = 0.16$. In addition, $r = s = 1$. The network wants to estimate a 2×1 unknown vector $w^\circ = [2; 3]$. In the simulation, we consider two situations. In one situation, the nodes use the average strategy, i.e., $a = 0.5$. In another situation, there is no cooperation and the nodes do not exchange information, i.e., $a = 0$. We observe that with the existence of the noise due to information exchange, the network without cooperation can possibly outperform the network with cooperation. We also show the theoretical steady-state network MSD (see (28) and (29)) in Fig. 2. The theoretical results match well with simulations.

We examine how the optimal combination weight and the corresponding MSD change as a function of the ratio t of σ_n^2 to σ_v^2, in Figs. 3 and 4. The values of a and the MSD are evaluated using (33) and (29), respectively. We compare two values of step size: $\mu = 0.01$ and $\mu = 0.05$. We observe that the nodes are sensitive to the noise $n_{k,i}$, especially for small step size, and a drops significantly when t is small. Using (33), we can find the value of t when $a = 0.25$ and obtain $t = -48.6$ dB for $\mu = 0.01$ and $t = -28.2$ dB for $\mu = 0.05$. In Fig. 4, the minimum MSD is compared to the MSD with the average strategy. The figure shows bifurcation. When the ratio t is small, these two networks have similar performance. However, after some critical value of t, the MSD

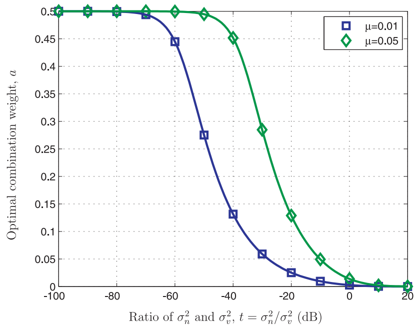

Fig. 3. Optimal combination weight, a, as a function of t.

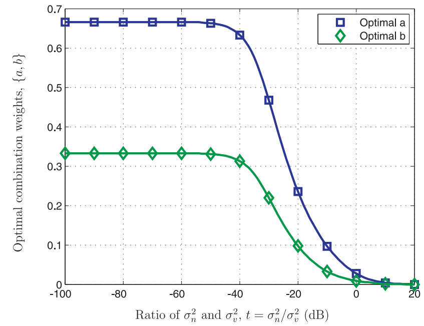

Fig. 5. Optimal combination weights, $\{a, b\}$, over t.

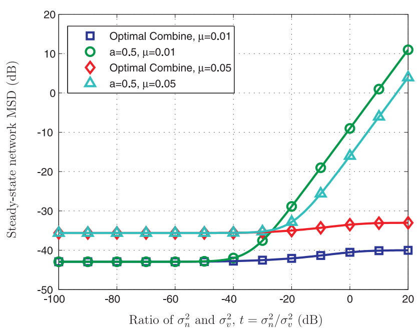

Fig. 4. Steady-state network MSD over t.

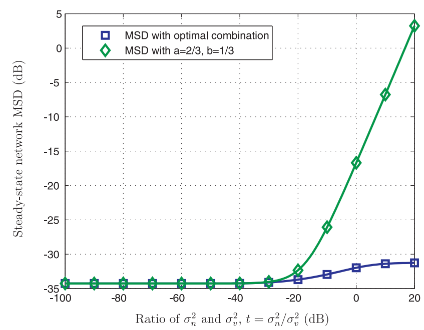

Fig. 6. Steady-state network MSD over t.

of the network using the averaging strategy increases linearly in t, whereas the optimal MSD increases slightly and saturates (converges to the network without cooperation). We also see that the critical value of t is smaller for a smaller step size. The results point to a trade-off in choosing the step-size parameter: smaller step-sizes achieve lower MSD but are more sensitive to link errors.

We also examine the situation when $r = 2$ and $s = 1$. That is, node 2 has better quality of measurement than node 1 while the noise levels due to information exchange are the same for both nodes. We show the optimal combination weights $\{a, b\}$ in Fig. 5 and the optimal MSD in Fig. 6. The optimal MSD is compared to the MSD with $\{a, b\} = \{2/3, 1/3\}$, which are the optimal values when there are no link errors. Similar curves are observed. It is interesting to note from Fig. 5 that node 1 places more weight on its own estimate (i.e., $a < 0.5$) when t is about -30 dB.

VI. Concluding Remarks

In this paper, we studied two-node adaptive networks with noisy links. The results show that, with noise from information exchange, a node should place more weight on its estimate in the combination step even if the node has worse quality of measurement than its neighbors. In addition, the results reveal a trade-off for choosing the step size as a compromise between the network MSD and sensitivity to link errors. Although the results are based on two-node networks due to space limitation, they can be extended to multiple nodes.

References

[1] S. Camazine, J. L. Deneubourg, N. R. Franks, J. Sneyd, G. Theraulaz, and E. Bonabeau, *Self-Organization in Biological Systems*. Princeton University Press, 2003.

[2] I. D. Couzin, "Collective cognition in animal groups," *Trends in Cognitive Sciences*, vol. 13, pp. 36–43, Jan. 2009.

[3] G. Turner and T. Pitcher, "Attack abatement: A model for group protection by combined avoidance and dilution," *American Naturalist*, vol. 128, pp. 228–240, 1986.

[4] M. Andersson and J. Wallander, "Kin selection and reciprocity in flight formation?" *Behavioral Ecology*, pp. 158–162, 2004.

[5] S. Janson, M. Middendorf, and M. Beekman, "Honeybee swarms: How do scouts guide a swarm of uninformed bees?" *Anim. Behav*, vol. 70, pp. 349–358, 2005.

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

[7] F. Cattivelli and A. H. Sayed, "Modeling bird flight formations using diffusion adaptation," *IEEE Trans. on Signal Processing*, vol. 59, no. 5, pp. 2038–2051, May 2011.

[8] J. Li and A. H. Sayed, "Modeling bee swarming behavior through diffusion adaptation with asymmetric information sharing," *EURASIP Journal on Advances in Signal Processing*, 2011.

[9] F. S. Cattivelli and A. H. Sayed, "Diffusion LMS strategies for distributed estimation," *IEEE Trans. on Signal Processing*, vol. 58, no. 3, pp. 1035–1048, Mar. 2010.

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

[11] X. Zhao and A. H. Sayed, "Performance limits of LMS-based adaptive networks," *Proc. IEEE ICASSP*, pp. 3768–3771, Prague, Czech Republic, May 2011.

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

[13] T. Y. Al-Naffouri and A. H. Sayed, "Transient analysis of data-normalized adaptive filters," *IEEE Trans. on Signal Processing*, vol. 51, no. 3, pp. 639–652, Mar. 2003.