Combination Weights for Diffusion Strategies With Imperfect Information Exchange

Xiaochuan Zhao and Ali H. Sayed
Department of Electrical Engineering
University of California, Los Angeles, CA 90095
Email: {xzhao, sayed}@ee.ucla.edu

Abstract—**Adaptive networks rely on in-network and collaborative processing among distributed agents to deliver enhanced performance in estimation and inference tasks. Information is exchanged among the nodes, usually over noisy links. This paper first investigates the mean-square performance of adaptive diffusion algorithms in the presence of various sources of imperfect information exchanges and quantization errors. Among other results, the analysis reveals that link noise over the regression data modifies the dynamics of the network evolution, and leads to biased estimates in steady-state. The analysis also reveals how the network mean-square performance is dependent on the combination weight matrices. We use these observations to show how the combination weights can be optimized and adapted. Simulation results illustrate the theoretical findings and match well with theory.**

Index Terms—**Diffusion adaptation, adaptive networks, imperfect information exchange, diffusion LMS, combination weights.**

I. Introduction

Diffusion strategies [1], [2] are distributed learning algorithms that can be used to solve distributed estimation and inference problems over networks effectively. In diffusion implementations, information is processed adaptively and locally at the nodes and then diffused in real-time across the network. Due to their robustness and scalability, diffusion strategies have been successfully applied to model self-organized and complex behavior in biological networks [3], [4] and to solve general optimization problems [5], [6].

In general, the information that is exchanged among nodes is subject to quantization errors and additive noise over the communication links. Studies on the effect of link noise on performance appear in [7]–[9] for consensus-type strategies and in [10]–[12] for diffusion-type strategies; the latter references simply extend the analysis of [1], [2] in a manner similar to what was done before in [13] while studying the performance of stand-alone filters under non-stationary conditions. In this paper, our objective is to go beyond these earlier works and to study the effect of link noise on network adaptation and learning by taking into account several additional sources of imperfection, and by considering more general algorithmic structures. The reason for this level of generality is because the analytical results will help reveal which noise sources influence the network performance more seriously, in what manner, and at what stage of the adaptation process. For example, the analysis will show that noise during the exchange of the regression data alters the learning dynamics of the network, and biases the weight estimates (see (41)). Noises related to the exchange of other pieces of information do not alter the dynamics of the network but contribute to the deterioration of the network performance. Our main objective is not simply to examine network performance under imperfect information exchange, but to show how to improve network performance when such imperfections exist. For this reason, we will use the analytical results to develop a method to select and adapt the combination weights in order to combat degradation caused by the various sources of information-exchange noise (see the algorithms in (57) and (59) further ahead). To arrive at these results, we allow for noisy exchanges during *each* of the three processing steps of a generalized adaptive diffusion algorithm (the two combination steps (3) and (5) and the adaptation step (4)). This level of generality enables us to examine how the three sets of combination coefficients $\{a_{1,lk}, c_{lk}, a_{2,lk}\}$ in (3)–(5) influence the propagation of all noises through the filter dynamics.

Notation: We use lowercase letters to denote vectors, uppercase letters for matrices, plain letters for deterministic variables, and boldface letters for random variables. We also use $(\cdot)^T$ to denote transposition, $(\cdot)^*$ for conjugate transposition, $(\cdot)^{-1}$ for matrix inversion, $\mathrm{Tr}(\cdot)$ for the trace of a matrix, and $\rho(\cdot)$ for the spectral radius of a matrix. Besides, we use $\otimes$ to denote Kronecker products, $\mathrm{vec}(A)$ to stack the columns of A on top of each other, and $\mathrm{unvec}(\cdot)$ to denote the inverse operation of $\mathrm{vec}(A)$. All vectors in our treatment are column vectors, with the exception of the regression vectors, $\boldsymbol{u}_{k,i}$, and the associated noise signals, $\boldsymbol{v}_{kl,i}^{(u)}$, which are taken to be row vectors for convenience of presentation.

II. Diffusion with Imperfect Information Exchange

We consider a connected network consisting of N nodes. Each node k collects scalar measurements $\boldsymbol{d}_k(i)$ and $1 \times M$ regression data vectors $\boldsymbol{u}_{k,i}$ over successive time instants $i \geq 0$. The measurements across all nodes are assumed to be related to an unknown $M \times 1$ vector w^o via a linear regression

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

model of the form [13]:

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

where $\boldsymbol{v}_k(i)$ denotes measurement or model noise with zero mean and variance $\sigma_{v,k}^2$. The vector w^o in (1) denotes the parameter of interest.

The nodes in the network would like to estimate w^o by seeking the solution that minimizes the global cost function:

$$\underset{w}{\text{minimize}} \sum_{k=1}^{N} \mathbb{E}|\boldsymbol{d}_k(i) - \boldsymbol{u}_{k,i} w|^2 \tag{2}$$

In previous works [1], [2], we introduced and studied several distributed strategies of the diffusion type for solving problems of the form (2) in an adaptive manner. They can be expressed in the following general form:

$$\boldsymbol{\phi}_{k,i-1} = \sum_{l \in \mathcal{N}_k} a_{1,lk} \boldsymbol{w}_{l,i-1} \tag{3}$$

$$\boldsymbol{\psi}_{k,i} = \boldsymbol{\phi}_{k,i-1} + \mu_k \sum_{l \in \mathcal{N}_k} c_{lk} \boldsymbol{u}_{l,i}^* [\boldsymbol{d}_l(i) - \boldsymbol{u}_{l,i} \boldsymbol{\phi}_{k,i-1}] \tag{4}$$

$$\boldsymbol{w}_{k,i} = \sum_{l \in \mathcal{N}_k} a_{2,lk} \boldsymbol{\psi}_{l,i} \tag{5}$$

where $a_{1,lk}$, c_{lk}, and $a_{2,lk}$ are nonnegative entries of $N \times N$ matrices A_1, C, and A_2, respectively. The coefficients $\{a_{1,lk}, c_{lk}, a_{2,lk}\}$ are zero whenever node l is not connected to node k, i.e., $l \notin \mathcal{N}_k$, where $\mathcal{N}_k$ denotes the neighborhood of node k. The so-called Combine-then-Adapt (CTA) [1] strategy can be recovered by setting $A_2 = I_N$, and the so-called Adapt-then-Combine (ATC) strategy [2] can be recovered by setting $A_1 = I_N$, where I_N denotes the $N \times N$ identity matrix. The matrices A_1, A_2, and C are required to be left or right-stochastic:

$$A_1^T \mathbb{1}_N = \mathbb{1}_N, \quad A_2^T \mathbb{1}_N = \mathbb{1}_N, \quad C \mathbb{1}_N = \mathbb{1}_N \tag{6}$$

where $\mathbb{1}_N$ denotes the $N \times 1$ all-one vector [2].

Each of the steps in (3)–(5) involves the sharing of information between node k and its neighbors. These exchange steps can all be subject to perturbations, such as additive noise and quantization errors. Our objective is to analyze the *aggregate* effect of these perturbations on general diffusion strategies of the type (3)–(5) and, more importantly, to propose choices for the scaling weights in order to enhance the mean-square performance of the network in the presence of these disturbances.

So let us model the data received by node k from its neighbor l as

$$\boldsymbol{w}_{lk,i-1} \triangleq \boldsymbol{w}_{l,i-1} + \boldsymbol{v}_{lk,i-1}^{(w)} \tag{7}$$

$$\boldsymbol{\psi}_{lk,i} \triangleq \boldsymbol{\psi}_{l,i} + \boldsymbol{v}_{lk,i}^{(\psi)} \tag{8}$$

$$\boldsymbol{d}_{lk}(i) \triangleq \boldsymbol{d}_l(i) + \boldsymbol{v}_{lk}^{(d)}(i) \tag{9}$$

$$\boldsymbol{u}_{lk,i} \triangleq \boldsymbol{u}_{l,i} + \boldsymbol{v}_{lk,i}^{(u)} \tag{10}$$

where $\boldsymbol{v}_{lk,i-1}^{(w)}$ and $\boldsymbol{v}_{lk,i}^{(\psi)}$ are $M \times 1$ noise signals, $\boldsymbol{v}_{lk,i}^{(u)}$ is a $1 \times M$ noise signal, and $\boldsymbol{v}_{lk}^{(d)}(i)$ is a scalar noise signal (see Fig. 1).

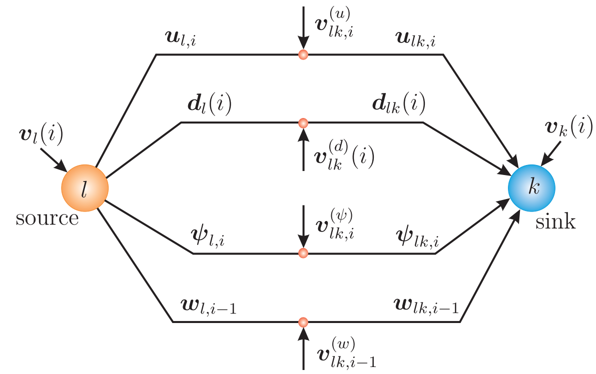


Fig. 1. Several additive noise sources perturb the exchange of information from node l to node k.

Observe that in (7)–(10), we are including several sources of information exchange noise. In comparison, references [10], [12] only considered the noise source $\boldsymbol{v}_{lk,i-1}^{(w)}$ in (7) and one set of scaling coefficients $\{a_{1,lk}\}$; the other coefficients were set to $c_{lk} = a_{2,lk} = 0$ for $l \neq k$ and $c_{kk} = a_{2,kk} = 1$. In other words, these references only considered (7) and the traditional CTA strategy without exchange of the data $\{\boldsymbol{d}_l(i), \boldsymbol{u}_{l,i}\}$.

To conduct the analysis, we introduce the following assumptions on the statistical properties of the measurement data and noise signals.

Assumption 1 (Statistical properties):

1) The regression data $\boldsymbol{u}_{k,i}$ are temporally and spatially independent and identically distributed (i.i.d.) random variables with zero mean and covariance matrix $R_{u,k} = \mathbb{E}\boldsymbol{u}_{k,i}^* \boldsymbol{u}_{k,i} > 0$.
2) The noise signals $\boldsymbol{v}_k(i)$, $\boldsymbol{v}_{lk,i}^{(w)}$, $\boldsymbol{v}_{lk}^{(d)}(i)$, $\boldsymbol{v}_{lk,i}^{(u)}$, and $\boldsymbol{v}_{lk,i}^{(\psi)}$ are temporally and spatially i.i.d. random variables with zero mean and covariances $\sigma_{v,k}^2$, $R_{v,lk}^{(w)}$, $\sigma_{v,lk}^2$, $R_{v,lk}^{(u)}$, and $R_{v,lk}^{(\psi)}$, respectively. In addition, $R_{v,lk}^{(w)}$, $\sigma_{v,lk}^2$, $R_{v,lk}^{(u)}$, and $R_{v,lk}^{(\psi)}$ are all zero if $l \notin \mathcal{N}_k$ or $l = k$.
3) The regression data $\boldsymbol{u}_{k,i}$, the model noise signals $\boldsymbol{v}_k(j_1)$, and the link noise signals $\boldsymbol{v}_{lk,j_2}^{(w)}$, $\boldsymbol{v}_{lk}^{(d)}(j_3)$, $\boldsymbol{v}_{lk,j_4}^{(u)}$, and $\boldsymbol{v}_{lk,j_5}^{(\psi)}$ are mutually-independent for all i and j_t, $t = 1, \ldots, 5$. ■

Using the perturbed data, the diffusion algorithm (3)–(5) becomes

$$\boldsymbol{\phi}_{k,i-1} = \sum_{l \in \mathcal{N}_k} a_{1,lk} \boldsymbol{w}_{l,i-1} + \boldsymbol{v}_{k,i-1}^{(w)} \tag{11}$$

$$\boldsymbol{\psi}_{k,i} = \boldsymbol{\phi}_{k,i-1} + \mu_k \sum_{l \in \mathcal{N}_k} c_{lk} \boldsymbol{u}_{lk,i}^* [\boldsymbol{d}_{lk}(i) - \boldsymbol{u}_{lk,i} \boldsymbol{\phi}_{k,i-1}] \tag{12}$$

$$\boldsymbol{w}_{k,i} = \sum_{l \in \mathcal{N}_k} a_{2,lk} \boldsymbol{\psi}_{l,i} + \boldsymbol{v}_{k,i}^{(\psi)} \tag{13}$$

where we are introducing the symbols $\boldsymbol{v}_{k,i-1}^{(w)}$ and $\boldsymbol{v}_{k,i}^{(\psi)}$ to denote the aggregate $M \times 1$ zero-mean noise signals defined over the neighborhood of node k:

$$\boldsymbol{v}_{k,i-1}^{(w)} \triangleq \sum_{l \in \mathcal{N}_k} a_{1,lk} \boldsymbol{v}_{lk,i-1}^{(w)}, \quad \boldsymbol{v}_{k,i}^{(\psi)} \triangleq \sum_{l \in \mathcal{N}_k} a_{2,lk} \boldsymbol{v}_{lk,i}^{(\psi)} \tag{14}$$

with covariance matrices

$$R_{v,k}^{(w)} \triangleq \sum_{l\in\mathcal{N}_k} a_{1,lk}^2 R_{v,lk}^{(w)}, \quad R_{v,k}^{(\psi)} \triangleq \sum_{l\in\mathcal{N}_k} a_{2,lk}^2 R_{v,lk}^{(\psi)} \tag{15}$$

We further introduce the following scalar zero-mean noise signal:

$$\boldsymbol{v}_{lk}(i) \triangleq \boldsymbol{v}_l(i) + \boldsymbol{v}_{lk}^{(d)}(i) - \boldsymbol{v}_{lk,i}^{(u)} w^o \tag{16}$$

Then, it is easy to verify that the noisy data $\{\boldsymbol{d}_{lk}(i), \boldsymbol{u}_{lk,i}\}$ are related via:

$$\boldsymbol{d}_{lk}(i) = \boldsymbol{u}_{lk,i} w^o + \boldsymbol{v}_{lk}(i) \tag{17}$$

In order to examine the evolution of the estimation error vectors across the network, we let

$$\widetilde{\boldsymbol{w}}_{k,i} \triangleq w^o - \boldsymbol{w}_{k,i} \tag{18}$$

and define

$$\boldsymbol{R}'_{k,i} \triangleq \sum_{l\in\mathcal{N}_k} c_{lk} \boldsymbol{u}_{lk,i}^* \boldsymbol{u}_{lk,i}, \quad \boldsymbol{z}_{k,i} \triangleq \sum_{l\in\mathcal{N}_k} c_{lk} \boldsymbol{u}_{lk,i}^* \boldsymbol{v}_{lk}(i) \tag{19}$$

We collect the various quantities across all nodes in the network into the following block vectors and matrices:

$$\boldsymbol{\mathcal{R}}'_i \triangleq \text{diag}\left\{\boldsymbol{R}'_{1,i}, \ldots, \boldsymbol{R}'_{N,i}\right\} \tag{20}$$

$$\boldsymbol{z}_i \triangleq \text{col}\{\boldsymbol{z}_{1,i}, \ldots, \boldsymbol{z}_{N,i}\} \tag{21}$$

$$\boldsymbol{v}_i^{(w)} \triangleq \text{col}\left\{\boldsymbol{v}_{1,i}^{(w)}, \ldots, \boldsymbol{v}_{N,i}^{(w)}\right\} \tag{22}$$

$$\boldsymbol{v}_i^{(\psi)} \triangleq \text{col}\left\{\boldsymbol{v}_{1,i}^{(\psi)}, \ldots, \boldsymbol{v}_{N,i}^{(\psi)}\right\} \tag{23}$$

$$\mathcal{M} \triangleq \text{diag}\{\mu_1 I_M, \ldots, \mu_N I_M\} \tag{24}$$

$$\widetilde{\boldsymbol{w}}_i \triangleq \text{col}\{\widetilde{\boldsymbol{w}}_{1,i}, \ldots, \widetilde{\boldsymbol{w}}_{N,i}\} \tag{25}$$

Then, some algebra will show that

$$\boxed{\begin{aligned} \widetilde{\boldsymbol{w}}_i = {} & \mathcal{A}_2^T \left(I_{NM} - \mathcal{M}\boldsymbol{\mathcal{R}}'_i\right) \mathcal{A}_1^T \widetilde{\boldsymbol{w}}_{i-1} - \mathcal{A}_2^T \mathcal{M} \boldsymbol{z}_i \\ & - \mathcal{A}_2^T \left(I_{NM} - \mathcal{M}\boldsymbol{\mathcal{R}}'_i\right) \boldsymbol{v}_{i-1}^{(w)} - \boldsymbol{v}_i^{(\psi)} \end{aligned}} \tag{26}$$

where

$$\mathcal{A}_1 \triangleq A_1 \otimes I_M, \quad \mathcal{C} \triangleq C \otimes I_M, \quad \mathcal{A}_2 \triangleq A_2 \otimes I_M \tag{27}$$

III. Steady-State Performance Analysis

We extend the analysis in [1], [2] by using the energy conservation argument [13] to examine the steady-state performance of the noisy diffusion strategy (11)–(13) in this section.

A. Variance Relation

The weighted variance relation for the error vector $\widetilde{\boldsymbol{w}}_i$ can be obtained from the error recursion (26) as:

$$\boxed{\begin{aligned} \mathbb{E}\,\|\widetilde{\boldsymbol{w}}_i\|_\Sigma^2 = {} & \mathbb{E}\|\widetilde{\boldsymbol{w}}_{i-1}\|_{\Sigma'}^2 + \mathbb{E}\|\mathcal{A}_2^T \mathcal{M} \boldsymbol{z}_i\|_\Sigma^2 \\ & -2\text{Re}\,\mathbb{E}[\boldsymbol{z}_i^* \mathcal{M}\mathcal{A}_2 \Sigma \mathcal{A}_2^T (I_{NM} - \mathcal{M}\boldsymbol{\mathcal{R}}'_i)\mathcal{A}_1^T \widetilde{\boldsymbol{w}}_{i-1}] \\ & +\mathbb{E}\|\mathcal{A}_2^T (I_{NM} - \mathcal{M}\boldsymbol{\mathcal{R}}'_i)\boldsymbol{v}_{i-1}^{(w)}\|_\Sigma^2 + \mathbb{E}\|\boldsymbol{v}_i^{(\psi)}\|_\Sigma^2 \end{aligned}} \tag{28}$$

where Σ is an arbitrary $NM \times NM$ positive semi-definite Hermitian matrix that we are free to choose. The matrix Σ' in (28), under Assumption 1, can be expressed as

$$\Sigma' = \mathcal{B}^* \Sigma \mathcal{B} + O(\mathcal{M}^2) \tag{29}$$

where

$$\mathcal{B} \triangleq \mathcal{A}_2^T \left(I_{NM} - \mathcal{M}\mathcal{R}'\right) \mathcal{A}_1^T \tag{30}$$

$$\mathcal{R}' \triangleq \mathbb{E}\boldsymbol{\mathcal{R}}'_i = \text{diag}\{R'_1, \ldots, R'_N\} \tag{31}$$

$$R'_k \triangleq \mathbb{E}\boldsymbol{R}'_{k,i} = \sum_{l\in\mathcal{N}_k} c_{lk} \left(R_{u,l} + R_{v,lk}^{(u)}\right) \tag{32}$$

Evaluating the term $O(\mathcal{M}^2)$ in (29) requires knowledge of higher-order statistics of the regression data and link noise that are not available under current assumptions. However, this term becomes negligible if we introduce a small step-size assumption.

Assumption 2 (Small step-sizes): The step-sizes are sufficiently small, i.e., $\mu_k \ll 1$. ■

It is shown in [14] that recursion (26) is stable in the mean and mean-square sense if the step-sizes are sufficiently small and satisfy:

$$\mu_k < \frac{2}{\rho\left(R'_k\right)}, \qquad k = 1, 2, \ldots, N \tag{33}$$

This condition is not only dependent on the regressors $\boldsymbol{u}_{k,i}$ but also on the link noise $\boldsymbol{u}_{lk,i}^{(u)}$. Then, in steady-state as $i \to \infty$, we get from (28)

$$\begin{aligned} \lim_{i\to\infty} \mathbb{E}\|\widetilde{\boldsymbol{w}}_i\|_{\Sigma-\Sigma'}^2 = {} & \text{Tr}(\mathcal{A}_2^T \mathcal{M}\mathcal{R}_z \mathcal{M}\mathcal{A}_2 \Sigma) + \text{Tr}(\mathcal{R}_v \Sigma) \\ & - \text{Tr}(\mathcal{Y}\Sigma) - \text{Tr}(\Sigma\mathcal{Y}^*) \end{aligned} \tag{34}$$

where, based on Assumptions 1 and 2,

$$\mathcal{R}_z \triangleq \mathbb{E}\,\boldsymbol{z}_i \boldsymbol{z}_i^* \tag{35}$$

$$g \triangleq \lim_{i\to\infty} \mathbb{E}\widetilde{\boldsymbol{w}}_{i-1} \tag{36}$$

$$\mathcal{Y} \triangleq \mathbb{E}[\mathcal{A}_2^T (I_{NM} - \mathcal{M}\boldsymbol{\mathcal{R}}'_i)\mathcal{A}_1^T g \boldsymbol{z}_i^* \mathcal{M}\mathcal{A}_2] \tag{37}$$

$$\mathcal{R}_v^{(w)} \triangleq \mathbb{E}\boldsymbol{v}_i^{(w)} \boldsymbol{v}_i^{(w)*} = \text{diag}\left\{R_{v,1}^{(w)}, \ldots, R_{v,N}^{(w)}\right\} \tag{38}$$

$$\mathcal{R}_v^{(\psi)} \triangleq \mathbb{E}\boldsymbol{v}_i^{(\psi)} \boldsymbol{v}_i^{(\psi)*} = \text{diag}\left\{R_{v,1}^{(\psi)}, \ldots, R_{v,N}^{(\psi)}\right\} \tag{39}$$

$$\begin{aligned} \mathcal{R}_v & \triangleq \mathbb{E}[\mathcal{A}_2^T (I_{NM} - \mathcal{M}\boldsymbol{\mathcal{R}}'_i)\boldsymbol{v}_{i-1}^{(w)} \boldsymbol{v}_{i-1}^{(w)*} \\ & \qquad \times (I_{NM} - \mathcal{M}\boldsymbol{\mathcal{R}}'_i)\mathcal{A}_2] + \mathbb{E}\boldsymbol{v}_i^{(\psi)} \boldsymbol{v}_i^{(\psi)*} \\ & = \mathcal{A}_2^T \mathcal{R}_v^{(w)} \mathcal{A}_2 + \mathcal{R}_v^{(\psi)} + O(\mathcal{M}) \\ & \approx \mathcal{A}_2^T \mathcal{R}_v^{(w)} \mathcal{A}_2 + \mathcal{R}_v^{(\psi)} \end{aligned} \tag{40}$$

The term g in (36) is the steady-state bias, which can be evaluated from (26) to be [14]:

$$\boxed{g = \left[I_{NM} - \mathcal{A}_2^T (I_{NM} - \mathcal{M}\mathcal{R}')\mathcal{A}_1^T\right]^{-1} \mathcal{A}_2^T \mathcal{M}\mathcal{R}_{v,c}^{(u)} (\mathbb{1}_N \otimes w^o)} \tag{41}$$

where the $NM \times NM$ matrix $\mathcal{R}_{v,c}^{(u)}$ collects all covariance matrices $\{R_{v,lk}^{(u)}\}$, $k, l = 1, \ldots, N$, weighted by the corresponding combination coefficients $\{c_{lk}\}$, such that its (k,l)th $M \times M$ submatrix is $c_{lk} R_{v,lk}^{(u)}$. It can be further verified from (29) and Assumption 2 that

$$\text{vec}(\Sigma') \approx (\mathcal{B}^T \otimes \mathcal{B}^*) \cdot \text{vec}(\Sigma) \tag{42}$$

Therefore, the steady-state weighted variance relation (34) becomes

$$\lim_{i\to\infty} \mathbb{E}\|\widetilde{\boldsymbol{w}}_i\|^2_{\text{unvec}[(I_{N^2M^2}-\mathcal{B}^T\otimes\mathcal{B}^*)\text{vec}(\Sigma)]} = \left[\text{vec}\left(\mathcal{A}_2^T\mathcal{M}\mathcal{R}_z\mathcal{M}\mathcal{A}_2+\mathcal{R}_v-\mathcal{Y}-\mathcal{Y}^*\right)\right]^*\text{vec}(\Sigma) \quad (43)$$

B. Network MSD and EMSE

The network MSD is defined as:

$$\overline{\text{MSD}} \triangleq \lim_{i\to\infty}\frac{1}{N}\sum_{k=1}^{N}\mathbb{E}\|\widetilde{\boldsymbol{w}}_{k,i}\|^2 \quad (44)$$

Since we are free to choose Σ, we select it as $(I_{N^2M^2}-\mathcal{B}^T\otimes\mathcal{B}^*)\text{vec}(\Sigma)=\text{vec}(I_{NM}/N)$. Then, expression (43) gives

$$\boxed{\overline{\text{MSD}} = \frac{1}{N}\left[\text{vec}\left(\mathcal{A}_2^T\mathcal{M}\mathcal{R}_z\mathcal{M}\mathcal{A}_2+\mathcal{R}_v-\mathcal{Y}-\mathcal{Y}^*\right)\right]^* \times\left(I_{N^2M^2}-\mathcal{B}^T\otimes\mathcal{B}^*\right)^{-1}\text{vec}(I_{NM})} \quad (45)$$

Similarly, if we instead select $(I_{N^2M^2}-\mathcal{B}^T\otimes\mathcal{B}^*)\text{vec}(\Sigma)=\text{vec}(\mathcal{R}_u/N)$, where

$$\mathcal{R}_u \triangleq \text{diag}\{R_{u,1},\ldots,R_{u,N}\} \quad (46)$$

then expression (43) would allow us to evaluate the network EMSE as:

$$\boxed{\overline{\text{EMSE}} = \frac{1}{N}\left[\text{vec}\left(\mathcal{A}_2^T\mathcal{M}\mathcal{R}_z\mathcal{M}\mathcal{A}_2+\mathcal{R}_v-\mathcal{Y}-\mathcal{Y}^*\right)\right]^* \times\left(I_{N^2M^2}-\mathcal{B}^T\otimes\mathcal{B}^*\right)^{-1}\text{vec}(\mathcal{R}_u)} \quad (47)$$

where the network EMSE is defined as follows:

$$\overline{\text{EMSE}} \triangleq \lim_{i\to\infty}\frac{1}{N}\sum_{k=1}^{N}\mathbb{E}|\boldsymbol{u}_{k,i}\widetilde{\boldsymbol{w}}_{k,i-1}|^2 \quad (48)$$

As seen from (36) and (41), the link noise introduced during the transmission of the regression data biases the estimators. We can now examine how the results simplify when there is no sharing of regression data among the nodes.

Assumption 3 (No sharing of regression data): Nodes do not share regression data within neighborhoods, i.e. assume $C = I_N$. ∎

By Assumptions 2 and 3, we simplify the network MSD expression (45) to:

$$\overline{\text{MSD}} = \frac{1}{N}\left[\text{vec}\left(\mathcal{A}_2^T\mathcal{M}\mathcal{S}\mathcal{M}\mathcal{A}_2+\mathcal{R}_v\right)\right]^* \times\left(I_{N^2M^2}-\mathcal{B}^T\otimes\mathcal{B}^*\right)^{-1}\text{vec}(I_{NM}) \quad (49)$$

where

$$\mathcal{S} \triangleq \text{diag}\left\{\sigma_{v,1}^2R_{u,1},\ldots,\sigma_{v,N}^2R_{u,N}\right\} \quad (50)$$

We can get the network MSD for the ATC algorithm by setting $A_1 = I$ and $A_2 = A$ (and similarly for the CTA algorithm by setting $A_1 = A$ and $A_2 = I$). Let us denote

$$\mathcal{A} \triangleq A\otimes I_M, \quad \mathcal{B}_{\text{atc}} \triangleq \mathcal{A}^T\left(I_{NM}-\mathcal{M}\mathcal{R}_u\right) \quad (51)$$

where $\mathcal{R}' = \mathcal{R}_u$ due to Assumption 3. Using the fact that $\mathcal{B}_{\text{atc}}$ is stable, we can arrive at another useful expression for the network MSD for ATC algorithms [14]:

$$\overline{\text{MSD}}_{\text{atc}} = \frac{1}{N}\sum_{j=0}^{\infty}\text{Tr}\left[\mathcal{B}_{\text{atc}}^j(\mathcal{A}^T\mathcal{M}\mathcal{S}\mathcal{M}\mathcal{A}+\mathcal{R}_v^{(\psi)})\mathcal{B}_{\text{atc}}^{*j}\right] \quad (52)$$

IV. Optimizing the Combination Matrices

Minimizing the MSD expression (52) for the ATC algorithm over left-stochastic matrices A is generally non-trivial. We pursue an approximate solution that relies on optimizing an upper bound and performs well in practice. It is shown in [14], [15] that the network MSD (52) is upper bounded by

$$\overline{\text{MSD}}_{\text{atc}} \le \frac{c^2}{N}\frac{\text{Tr}\left(\mathcal{A}^T\mathcal{M}\mathcal{S}\mathcal{M}\mathcal{A}+\mathcal{R}_v^{(\psi)}\right)}{1-\rho\left(I_{NM}-\mathcal{M}\mathcal{R}_u\right)^2} \quad (53)$$

where the combination matrix $\mathcal{A}$ appears only in the numerator. This result motivates us to consider instead the problem of minimizing the upper bound, namely,

$$\begin{array}{ll}\underset{A}{\text{minimize}} & \text{Tr}\left(\mathcal{A}^T\mathcal{M}\mathcal{S}\mathcal{M}\mathcal{A}+\mathcal{R}_v^{(\psi)}\right)\\ \text{subject to} & A^T\mathbb{1}=\mathbb{1},\ \ a_{lk}\ge 0,\ \ a_{lk}=0 \text{ if } l\notin\mathcal{N}_k\end{array} \quad (54)$$

Using (39), problem (54) can be decoupled into N separate optimization problems of the form:

$$\boxed{\begin{array}{ll}\underset{\{a_{lk},\,l\in\mathcal{N}_k\}}{\text{minimize}} & \displaystyle\sum_{l\in\mathcal{N}_k}a_{lk}^2\left[\mu_l^2\sigma_{v,l}^2\text{Tr}\left(R_{u,l}\right)+\text{Tr}\left(R_{v,lk}^{(\psi)}\right)\right]\\ \text{subject to} & \displaystyle\sum_{l\in\mathcal{N}_k}a_{lk}=1,\ a_{lk}\ge 0,\ a_{lk}=0 \text{ if } l\notin\mathcal{N}_k\end{array}} \quad (55)$$

where $k = 1, 2, \ldots, N$. With each node l in the neighborhood $\mathcal{N}_k$, we associate the following nonnegative variance product:

$$\gamma_{lk}^2 \triangleq \begin{cases}\mu_k^2\sigma_{v,k}^2\text{Tr}\left(R_{u,k}\right), & l=k\\ \mu_l^2\sigma_{v,l}^2\text{Tr}\left(R_{u,l}\right)+\text{Tr}\left(R_{v,lk}^{(\psi)}\right), & l\in\mathcal{N}_k\backslash\{k\}\end{cases} \quad (56)$$

This metric incorporates information about the noise covariances $\{R_{v,lk}^{(\psi)}\}$. Then, the solution of (55) is given by

$$\boxed{a_{lk} = \begin{cases}\dfrac{\gamma_{lk}^{-2}}{\sum_{l\in\mathcal{N}_k}\gamma_{lk}^{-2}}, & \text{if } l\in\mathcal{N}_k\\ 0, & \text{otherwise}\end{cases}} \quad \text{(relative variance rule)} \quad (57)$$

We refer to this combination rule as the relative-variance combination rule; it is an extension of the rule devised in [15] to the case of noisy information exchanges. Minimizing the upper bound of the network EMSE for the ATC algorithm over left-stochastic matrices A leads to the same solution (57). Using the same argument, we can also show that the same result minimizes an upper bound on the network MSD or EMSE for the CTA algorithm.

The relative variance combination rule (57) relies on knowledge of the quantities $\{\sigma_{v,l}^2, \text{Tr}(R_{u,l}), \text{Tr}(R_{v,lk}^{(\psi)})\}$, which in general are not available. Therefore, we shall propose an adaptive combination rule. It is shown in [14], [15] that under Assumption 2, $\{\gamma_{lk}^2\}$ can be estimated through the recursion:

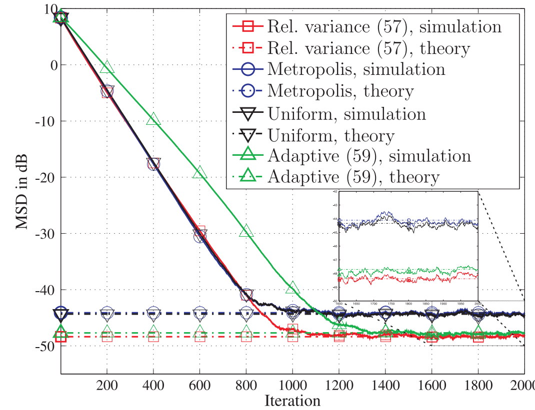
$$\boxed{\widehat{\gamma}_{lk}^2(i) = (1-\nu_k)\widehat{\gamma}_{lk}^2(i-1)+\nu_k\|\boldsymbol{\psi}_{lk,i}-\boldsymbol{w}_{k,i-1}\|^2} \tag{58}$$

where ν_k is a positive coefficient smaller than one. Indeed, it can be verified that the estimator $\widehat{\gamma}_{lk}^2(i)$ converges on average to the desired variance product γ_{lk}^2 with an error on the order of $O(\mu_l)+O(\mu_k)$, which is negligible under Assumption 2. In this way, we arrive at the adaptive combination rule:

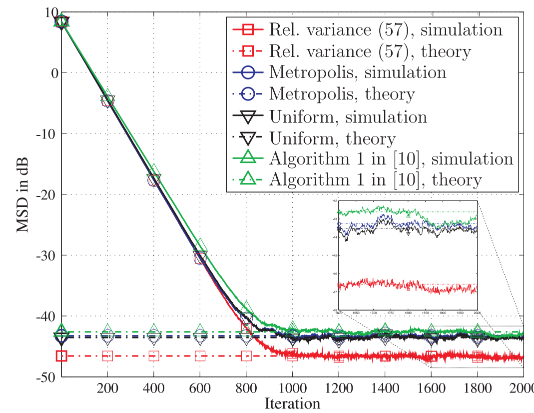
$$\boxed{a_{lk}(i) = \begin{cases} \dfrac{\widehat{\gamma}_{lk}^{-2}(i)}{\sum_{l\in\mathcal{N}_k}\widehat{\gamma}_{lk}^{-2}(i)}, & \text{if } l\in\mathcal{N}_k \\ 0, & \text{otherwise} \end{cases}} \tag{59}$$

V. Simulation Results and Conclusions

We consider a connected network with $N = 20$ nodes with a randomly generated topology. The unknown complex parameter w^o of length $M = 2$ is randomly generated. We adopt a uniform step-size, $\mu = 0.01$, and uniformly white Gaussian regression data, $R_{u,k} = \sigma_{u,k}^2 I_M$ where $\sigma_{u,k}^2 = 0.5$ for $k \in \{1, 2, \ldots, N/2\}$ and $\sigma_{u,k}^2 = 1$ for $k \in \{N/2+1, N/2+2, \ldots, N\}$. We also use white Gaussian link noise signals such that $R_{v,lk}^{(w)} = \sigma_{w,lk}^2 I_M$, $R_{v,lk}^{(u)} = \sigma_{u,lk}^2 I_M$, and $R_{v,lk}^{(\psi)} = \sigma_{\psi,lk}^2 I_M$. All noise variances, $\{\sigma_{v,k}^2, \sigma_{w,lk}^2, \sigma_{u,lk}^2, \sigma_{v,lk}^2, \sigma_{\psi,lk}^2\}$, are randomly generated. The signal-to-noise ratio across the network, $\frac{1}{N}\sum_{k=1}^{N}\mathrm{Tr}(R_{u,k})/\sigma_{v,k}^2$, is set to approximately 25 dB. The average power of each type of link noise across the network is 35 dB less than that of the model noise. We examine the traditional CTA and ATC algorithms without sharing data among nodes (i.e., $C = I_N$), under various combination rules: (i) the relative variance rule in (57), (ii) the Metropolis rule in [2], (iii) the uniform weighting rule [2], and (iv) the adaptive rule in (59). We plot the network MSD learning curves for CTA algorithms ($A_1 = A$ and $A_2 = I_N$) in Fig. 2b and for ATC algorithms ($A_1 = I_N$ and $A_2 = A$) in Fig. 2a, respectively, by averaging over 50 experiments. We also plot the theoretical result (49) in the same figures. We see that the relative variance rule makes diffusion algorithms achieve lower MSD levels at steady-state compared to the metropolis and uniform rules, as well as the algorithm from [10] (which requires knowledge of the noise variances). In addition, the adaptive rule attains MSD levels that are only slightly larger than those of the relative variance rule, although, as expected, it converges slower due to the additional learning step.



(a) Network MSD curves for ATC algorithms



(b) Network MSD curves for CTA algorithms

Fig. 2. Simulated network MSD curves for diffusion algorithms with noisy information exchange.

References

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

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

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

[4] S-Y. Tu and A. H. Sayed, "Mobile adaptive networks," *IEEE J. Sel. Top. Signal Process.*, vol. 5, no. 4, pp. 649–664, Aug. 2011.

[5] J. Chen and A. H. Sayed, "Diffusion adaptation strategies for distributed optimization and learning over networks," arXiv:1111.0034v1 [math.OC], Oct. 2011.

[6] J. Chen and A. H. Sayed, "Performance of diffusion adaptation for collaborative optimization," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, Kyoto, Japan, Mar. 2012.

[7] S. Kar and J. M. F. Moura, "Sensor networks with random links: Topology design for distributed consensus," *IEEE Trans. Signal Process.*, vol. 56, no. 7, pp. 3315–3326, July 2008.

[8] S. Kar and J. M. F. Moura, "Distributed consensus algorithms in sensor networks: Link failures and channel noise," *IEEE Trans. Signal Process.*, vol. 57, no. 1, pp. 355–369, Jan. 2009.

[9] K. Srivastava and A. Nedic, "Distributed asynchronous constrained stochastic optimization," *IEEE J. Sel. Top. Signal Process.*, vol. 5, no. 4, pp. 772–790, Aug. 2011.

[10] R. Abdolee and B. Champagne, "Diffusion LMS algorithms for sensor networks over non-ideal inter-sensor wireless channels," in *Proc. IEEE Int. Conf. Dist. Comput. Sensor Systems (DCOSS)*, Barcelona, Spain, June 2011, pp. 1–6.

[11] S-Y. Tu and A. H. Sayed, "Adaptive networks with noisy links," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Houston, TX, Dec. 2011.

[12] A. Khalili, M. A. Tinati, A. Rastegarnia, and J. A. Chambers, "Steady-state analysis of diffusion LMS adaptive networks with noisy links," *IEEE Trans. Signal Process.*, vol. 60, no. 2, pp. 974–979, Feb. 2012.

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

[14] X. Zhao and A. H. Sayed, "Diffusion adaptation over netowrks under imperfect information exchange and non-stationary data," arXiv:1112.6212v1 [math.OC], Dec. 2011.

[15] S-Y. Tu and A. H. Sayed, "Optimal combination rules for adaptation and learning over netowrks," in *Proc. IEEE Int. Workshop Comput. Advances Multi-Sensor Adapt. Process. (CAMSAP)*, San Juan, Puerto Rico, Dec. 2011.