PROBABILITY DISTRIBUTION OF STEADY-STATE ERRORS AND ADAPTATION OVER NETWORKS

Xiaochuan Zhao and Ali H. Sayed

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

ABSTRACT

In this work, we derive a near-optimal combination rule for adaptation over networks. To do so, we first establish a useful result pertaining to the steady-state distribution of the estimator of an LMS filter. Specifically, under small step-sizes and some conditions on the data, we show that the steady-state estimator is approximately Gaussian and provide an expression for its covariance matrix. The result is subsequently used to show that the maximum ratio combining rule over networks, which is used to combine the estimators across neighbors within a network, is near optimal in the minimum variance unbiased sense. The result suggests a rule for combining the estimators within neighborhoods that can lead to improved mean-square error performance.

Index Terms— LMS adaptation, steady-state distribution, steady-state error, adaptation over networks.

1. INTRODUCTION

Consider a collection of N LMS-based filters for estimating a weight vector w^o of size M. Let $\boldsymbol{w}_{k,\infty}$, $k = 1, \ldots, N$, denote the steady-state estimator by each filter, assuming the step-sizes are small enough to ensure filter convergence in the mean-square sense [1]. Such collections of filters arise in the study of adaptive networks, with nodes spread over a spatial domain [2] [3] [4]. The N filters could refer to nodes within any particular neighborhood in the network – see Fig. 1. A question of interest is to investigate how to combine the estimators provided by the individual nodes to obtain an estimator $\boldsymbol{w}^c$ that can outperform the estimators $\{\boldsymbol{w}_{k,\infty}\}$ in the mean-square-error sense. To address this question, we first take a step-back and derive a result for the classical LMS filter from a new perspective. This viewpoint will allow us to obtain an insight that facilitates the computation of $\boldsymbol{w}^c$ and the assessment of its performance. Specifically, we will first show that, for small step-sizes, and under some assumptions on the data, the steady-state estimator $\boldsymbol{w}_{k,\infty}$ for the kth filter is approximately Gaussian distributed with mean w^o and covariance matrix $\mu_k \sigma_{v,k}^2 (2I - \mu R_{u,k})^{-1}$, where μ_k is the step-size, $\sigma_{v,k}^2$ is the variance of noise at node k, and $R_{u,k}$ is the covariance matrix of the regression data also at node k. The main observation here is the determination that $w_{k,\infty}$ is approximately *Gaussian* distributed. Under this Gaussianity condition, we will then be able to derive an expression for the (near) optimal combination $\boldsymbol{w}^c$. We use the term "near-optimal" as opposed to "optimal" to highlight the fact that we are establishing *approximate* Gaussianity in steady-state as opposed to exact Gaussianity. As the simulations will indicate, the approximate Gaussianity result is well justified. Then, by appealing to the Gauss-Markov Theorem [1], we provide an expression for the minimum-variance unbiased estimator (MVUE) $\boldsymbol{w}^c$ (see (40) further ahead). In the special case when $R_{u,k} = R_u$ and $\mu_k = \mu$ across the nodes, our result (40) will simplify in (43) to the conclusion that the optimal $\boldsymbol{w}^c$ can be obtained by means of the maximal ratio combining (MRC) rule applied to the individual estimators $\{\boldsymbol{w}_{k,\infty}\}$, namely,

$$\boldsymbol{w}^c \triangleq \frac{1}{\sum_{l=1}^{N} \sigma_{v,l}^{-2}} \cdot \left(\sum_{k=1}^{N} \frac{1}{\sigma_{v,k}^2} \boldsymbol{w}_{k,\infty} \right) \tag{1}$$

Given the Gaussianity property, as established by Theorem 3.4, the discussion in the paper therefore establishes that the linear combination rules (1) and (40) are (near-) optimal in the mean-square sense over the class of both linear and nonlinear estimators for w^o (and not only over the class of linear estimators).

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

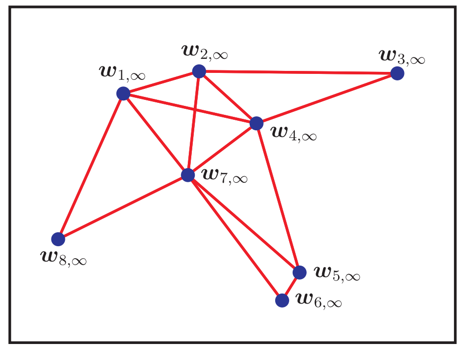

Fig. 1. A neighborhood with adaptive nodes estimating a parameter vector w^o.

2. DATA MODEL AND ASSUMPTIONS

We consider a linear regression model of the form:

$$\boldsymbol{d}(i) = \boldsymbol{u}_i w^o + \boldsymbol{v}(i) \tag{2}$$

where w^o is the unknown vector of size $M \times 1$, $\boldsymbol{u}_i$ is the regression vector of size $1 \times M$ at time i, and $\boldsymbol{v}(i)$ is noise also at time i. All random variables are assumed to have zero means. We focus initially on a single node running an LMS filter. We are interested in estimating the unknown parameter w^o. We use the least-mean squares (LMS) algorithm [1] to update the weight estimator as:

$$\boldsymbol{w}_i = \boldsymbol{w}_{i-1} + \mu \boldsymbol{u}_i^* (\boldsymbol{d}(i) - \boldsymbol{u}_i \boldsymbol{w}_{i-1}) \tag{3}$$

978-1-4577-0570-0/11/$26.00 ©2011 IEEE

The error vector at time i is defined as

$$\widetilde{\boldsymbol{w}}_i \triangleq w^o - \boldsymbol{w}_i \tag{4}$$

Substituting (2) and (3) into (4) yields the recursion

$$\widetilde{\boldsymbol{w}}_i = (I - \mu \boldsymbol{u}_i^* \boldsymbol{u}_i)\widetilde{\boldsymbol{w}}_{i-1} - \mu \boldsymbol{u}_i^* \boldsymbol{v}(i) \tag{5}$$

As is customary in the literature on LMS adaptation, we adopt the following assumptions:

Assumption 2.1 *The regression data $\boldsymbol{u}_i$ are independent and identically distributed (i.i.d.) real Gaussian random variables with zero mean and covariance matrix R_u.*

Assumption 2.2 *The noise signals $\boldsymbol{v}(i)$ are i.i.d. real Gaussian random variables with zero mean and variances σ_v^2.*

Assumption 2.3 *The regression data $\boldsymbol{u}_i$ and the noise signals $\boldsymbol{v}(j)$ are independent of each other for all i and j.*

Assumption 2.4 *The step size μ is sufficiently small such that*

$$\mu \mathsf{E}\|\boldsymbol{u}_i\|^2 = \mu \mathrm{Tr}(R_u) \ll 1 \tag{6}$$

to guarantee the mean-square convergence of LMS.

It was argued in [5] that $\widetilde{\boldsymbol{w}}_i$ converges in distribution; it was further noted that the limiting distribution is difficult to characterize. Based on the data model (2) and Assumptions 2.1-2.4, we will actually establish that the probability density function (pdf) of the estimator $\boldsymbol{w}_i$ at steady-state, i.e., $\boldsymbol{w}_\infty$, is well approximated by a Gaussian distribution.

3. PDF OF THE WEIGHT ERROR VECTOR

Using a standard averaging theory approximation for small step-sizes [6], we replace the instantaneous product $\boldsymbol{u}_i^*\boldsymbol{u}_i$ by its expectation R_u and write (5) as

$$\widetilde{\boldsymbol{w}}_i \approx (I - \mu R_u)\widetilde{\boldsymbol{w}}_{i-1} - \mu \boldsymbol{u}_i^* \boldsymbol{v}(i) \tag{7}$$

We further introduce the eigenvalue decomposition of R_u, say,

$$R_u = U \Lambda U^* \tag{8}$$

where U is an $M \times M$ unitary matrix and Λ is an $M \times M$ positive-definite diagonal matrix. Introduce the transformed variables:

$$\overline{w}^o = U^* w^o, \;\; \overline{\boldsymbol{w}}_i = U^* \widetilde{\boldsymbol{w}}_i, \;\; \overline{\boldsymbol{u}}_i = \boldsymbol{u}_i U \tag{9}$$

Then (7) can be transformed into

$$\overline{\boldsymbol{w}}_i \approx (I - \mu\Lambda)\overline{\boldsymbol{w}}_{i-1} - \mu \overline{\boldsymbol{u}}_i^* \boldsymbol{v}(i) \tag{10}$$

Assuming $\boldsymbol{w}_{-1} = 0$, the initial error vector becomes

$$\widetilde{\boldsymbol{w}}_{-1} = w^o - \boldsymbol{w}_{-1} = w^o \tag{11}$$

so that

$$\overline{\boldsymbol{w}}_{-1} = U^* \widetilde{\boldsymbol{w}}_{-1} = U^* w^o = \overline{w}^o \tag{12}$$

Then, after n iterations, we find from (10) that

$$\overline{\boldsymbol{w}}_n \approx (I - \mu\Lambda)^{n+1}\overline{w}^o - \mu \sum_{k=0}^{n} (I - \mu\Lambda)^{n-k} \overline{\boldsymbol{u}}_k^* \boldsymbol{v}(k) \tag{13}$$

Taking expectations of both sides and recalling that $\overline{\boldsymbol{u}}_k$ and $\boldsymbol{v}(k)$ are independent of each other, we obtain

$$\mathsf{E}\overline{\boldsymbol{w}}_n = (I - \mu\Lambda)^{n+1}\overline{w}^o \tag{14}$$

Actually, expression (14) is an exact relation (and not an approximation) under Assumptions 2.1-2.3. Indeed, if we take expectations of both sides of (5), we get

$$\begin{aligned} \mathsf{E}\widetilde{\boldsymbol{w}}_i &= (I - \mu R_u)\mathsf{E}\widetilde{\boldsymbol{w}}_{i-1} \\ &= (I - \mu R_u)^{i+1}\mathsf{E}\widetilde{\boldsymbol{w}}_{-1} \end{aligned} \tag{15}$$

where the first equality follows from Assumptions 2.1-2.3. Expression (15) can be easily transformed into (14). Now, since (14) is decaying exponentially, we conclude that $\boldsymbol{w}_n$ is an asymptotically unbiased estimator for w^o. This result is well-known in the adaptive filtering literature [1].

We now proceed to introduce a centered error vector in order to investigate the statistical distribution of $\overline{\boldsymbol{w}}_n$ in steady-state. Let

$$\begin{aligned} \overline{\overline{\boldsymbol{w}}}_n &\triangleq \overline{\boldsymbol{w}}_n - \mathsf{E}\overline{\boldsymbol{w}}_n \\ &= \overline{\boldsymbol{w}}_n - (I - \mu\Lambda)^{n+1}\overline{w}^o \\ &\approx -\mu \sum_{k=0}^{n} (I - \mu\Lambda)^{n-k} \overline{\boldsymbol{u}}_k^* \boldsymbol{v}(k) \end{aligned} \tag{16}$$

Because $I - \mu\Lambda$ is diagonal, expression (16) can be expressed element-wise as

$$\begin{aligned} \overline{\overline{\boldsymbol{w}}}(n,m) &\approx -\mu \sum_{k=0}^{n} (1 - \mu\lambda_m)^{n-k} \overline{\boldsymbol{u}}(k,m)^* \boldsymbol{v}(k) \\ &= -\mu \sum_{k=0}^{n} (1 - \mu\lambda_m)^{k} \overline{\boldsymbol{u}}(n-k,m)^* \boldsymbol{v}(n-k) \end{aligned} \tag{17}$$

where $\overline{\overline{\boldsymbol{w}}}(k,m)$ and $\overline{\boldsymbol{u}}(k,m)$ denote the mth entries of $\overline{\overline{\boldsymbol{w}}}_k$ and $\overline{\boldsymbol{u}}_k$, respectively, $m = 1, \ldots, M$. Note that (17) is a weighted sum of a set of i.i.d. random variables $\{\overline{\boldsymbol{u}}(k,m)^*\boldsymbol{v}(k)\}$, $k = 0, \ldots, n$. In the sequel we show that $\overline{\overline{\boldsymbol{w}}}(k,m)$ can be well approximated by a Gaussian random variable. We first establish three auxiliary lemmas. The proofs for the first two lemmas are omitted due to limited space.

Lemma 3.1 *Let $\mathbf{u}$ and $\mathbf{v}$ denote two mutually-independent Gaussian random variables, i.e., $\mathbf{u} \sim \mathbb{N}(0, \sigma_u^2)$ and $\mathbf{v} \sim \mathbb{N}(0, \sigma_v^2)$. Then, the pdf of the product $\mathbf{x} = \mathbf{u}\mathbf{v}$ is [7]*

$$p_{\mathbf{x}}(x) = \frac{1}{\pi \sigma_u \sigma_v} K_0\left(\frac{|x|}{\sigma_u \sigma_v}\right) \tag{18}$$

where $K_0(\cdot)$ is the zeroth-order modified Bessel function of the second kind [8], defined as

$$K_0(x) = \int_0^\infty \cos(x \sinh(t))dt = \int_0^\infty \frac{\cos(xt)}{\sqrt{t^2+1}}dt \tag{19}$$

Moreover, the characteristic function of $\mathbf{x}$ is

$$\phi_{\mathbf{x}}(\xi) = \frac{1}{\sqrt{1 + \sigma_u^2 \sigma_v^2 \xi^2}} \tag{20}$$

Lemma 3.2 *Let $\{\mathbf{x}_k\}$, $k = 0, \ldots, n$, be a set of i.i.d. random variables, each distributed according to the pdf (18). The normalized weighted sum*

$$\mathbf{y}_n = \frac{1}{\sigma_u \sigma_v} \sqrt{\frac{1-\rho^2}{1-\rho^{n+1}}} \sum_{k=0}^{n} \rho^k \mathbf{x}_k \tag{21}$$

where $0 < \rho < 1$ is a constant, has the characteristic function

$$\phi_{\mathbf{y}_n}(\xi) = \prod_{k=0}^{n} \frac{1}{\sqrt{1+\rho^{2k}\frac{1-\rho^2}{1-\rho^{n+1}}\xi^2}} \tag{22}$$

When n is asymptotically infinite, (22) becomes

$$\phi_{\mathbf{y}_\infty}(\xi) \triangleq \lim_{n\to\infty} \phi_{\mathbf{y}_n}(\xi) = \left((\rho^2-1)\xi^2;\rho^2\right)_\infty^{-1/2} \tag{23}$$

where the notation $(a;q)_\infty$ is the q-analog of the Pochhammer symbol (q-Pochhammer symbol, or q-series) [8], defined as

$$(a;q)_\infty \triangleq \prod_{k=0}^{\infty}(1-aq^k) \tag{24}$$

Lemma 3.3 *The function $\phi_{\mathbf{y}_\infty}(\xi)$ in (23) is well approximated by the characteristic function of the standard Gaussian random variable, which is given by*

$$\phi_o(\xi) \triangleq e^{-\xi^2/2} \tag{25}$$

when $1-\rho \ll 1$. This fact implies that $\mathbf{y}_\infty$ is approximately distributed as a standard Gaussian random variable, i.e., $\mathbf{y}_\infty \sim \mathbb{N}(0,1)$.

Proof First, by (23), the natural logarithm of $\phi_{\mathbf{y}_\infty}(\xi)$ is

$$\ln\left(\phi_{\mathbf{y}_\infty}(\xi)\right) = -\frac{1}{2}\sum_{k=0}^{\infty}\ln\left(1+\rho^{2k}(1-\rho^2)\xi^2\right) \tag{26}$$

Expanding (26) by using the Maclaurin Series expansion of the natural logarithm function $\ln(1+x)$ yields

$$\ln\left(\phi_{\mathbf{y}_\infty}(\xi)\right) = -\frac{1}{2}\sum_{k=0}^{\infty}\rho^{2k}(1-\rho^2)\xi^2 + O\left((1-\rho^2)^2\right) \tag{27}$$

Let $\epsilon \triangleq 1-\rho \ll 1$. Then, $1-\rho^2 = 2\epsilon - \epsilon^2 \approx 2\epsilon \ll 1$. So, (27) can be expressed as

$$\ln\left(\phi_{\mathbf{y}_\infty}(\xi)\right) = -\frac{(1-\rho^2)\xi^2}{2}\sum_{k=0}^{\infty}\rho^{2k} + o(\epsilon) \approx -\frac{\xi^2}{2}$$

which completes the proof. ∎

Theorem 3.4 *At steady-state, the random variable $\overline{\overline{\boldsymbol{w}}}(k,m)$ in (17) is well approximated by a Gaussian random variable with zero mean and variance $\frac{\mu\sigma_v^2}{2-\mu\lambda_m}$.*

Proof First, because $\overline{\boldsymbol{u}}(k,m)^*$ and $\boldsymbol{v}(k)$ are mutually-independent zero-mean real Gaussian random variables, by Lemma 3.1, the pdf of $\mathbf{x}_{k,m} \triangleq \overline{\boldsymbol{u}}(n-k,m)^*\boldsymbol{v}(n-k)$ is

$$p_{\mathbf{x}_{k,m}}(x) = \frac{1}{\pi\sqrt{\lambda_m\sigma_v^2}}K_0\left(\frac{|x|}{\sigma_u\sigma_v}\right) \tag{28}$$

At steady-state and using Lemmas 3.2 and 3.3, the normalized weighted sum

$$\mathbf{y}_{\infty,m} = \sqrt{\frac{1-\rho^2}{\lambda_m\sigma_v^2}}\sum_{k=0}^{\infty}\rho^k\mathbf{x}_{k,m} \tag{29}$$

with $\rho = 1-\mu\lambda_m \in (0,1)$ has a characteristic function that is well approximated by that of the standard Gaussian random variable, when $\mu\lambda_m \le \mu\mathrm{Tr}(R_u) \ll 1$. Because the characteristic function of a random variable defines its probability distribution, we conclude that $\mathbf{y}_{\infty,m}$ is well approximated by the standard Gaussian random variable. Moreover, we can make the identification:

$$\overline{\overline{\boldsymbol{w}}}(\infty,m) \approx -\mu\sqrt{\frac{\lambda_m\sigma_v^2}{1-\rho^2}}\mathbf{y}_{\infty,m} \tag{30}$$

which implies that, approximately,

$$\overline{\overline{\boldsymbol{w}}}(\infty,m) \sim \mathbb{N}\left(0,\frac{\mu\sigma_v^2}{2-\mu\lambda_m}\right) \tag{31}$$

as desired. ∎

Remark: The classical Central Limit Theorem (CLT) [7] cannot be applied directly to (17) to conclude that the entries $\overline{\overline{\boldsymbol{w}}}(n,m)$ satisfy a Gaussian distribution. This is because the terms that appear inside the summation in (17) do not satisfy the conditions that are usually required by CLT. In particular, Lemma 3.2 indicates that the sum does not converge to an actual Gaussian distribution in steady-state. Only when $\rho = 1$, $\overline{\overline{\boldsymbol{w}}}(n,m)$ is asymptotically Gaussian in n, which can also be concluded by CLT.

Based on this theorem, we can get a corollary on the limiting distribution of the estimate $\boldsymbol{w}_i$ when i goes to infinity.

Corollary 3.5 *At steady-state, the estimate $\boldsymbol{w}_\infty$ is approximately distributed as a Gaussian random vector, i.e.,*

$$\boldsymbol{w}_\infty \sim \mathbb{N}\left(w^o, \mu\sigma_v^2(2I-\mu R_u)^{-1}\right) \tag{32}$$

Proof According to (14), we already know that

$$\mathsf{E}\overline{\boldsymbol{w}}_\infty \triangleq \lim_{n\to\infty}\mathsf{E}\overline{\boldsymbol{w}}_n = \lim_{n\to\infty}(I-\mu\Lambda)^{n+1}\overline{w}^o = 0 \tag{33}$$

Then, some algebra shows that (32) follows from (16) and (17), and from Theorem 3.4. ∎

With Assumption 2.4, the mean-square-deviation (MSD) can be approximated as

$$\mathrm{MSD} \triangleq \mathsf{E}\|\widetilde{\boldsymbol{w}}_\infty\|^2 = \mathsf{E}\|\overline{\boldsymbol{w}}_\infty\|^2 \approx \sum_{m=1}^{M}\frac{\mu\sigma_v^2}{2-\mu\lambda_m} \approx \frac{\mu M\sigma_v^2}{2}$$

which coincides with the small-step-size approximation for the MSD of LMS according to (23.58) in [1].

4. NEAR-OPTIMAL COMBINATION OVER NETWORKS

Let us return to a neighborhood consisting of N nodes, as indicated by Fig. 1. According to Corollary 3.5, the steady-state estimator for the kth node, denoted by $\boldsymbol{w}_{k,\infty}$, is approximately distributed as

$$\boldsymbol{w}_{k,\infty} \sim \mathbb{N}(w^o, P_k) \tag{34}$$

where $P_k = \mu_k\sigma_{v,k}^2(2I-\mu_k R_{u,k})^{-1}$. Define

$$\boldsymbol{\mathcal{W}}_\infty \triangleq \mathrm{col}(\boldsymbol{w}_{1,\infty},\ldots,\boldsymbol{w}_{N,\infty}) \tag{35}$$

$$\mathcal{P} \triangleq \mathrm{diag}(P_1,\ldots,P_N) \tag{36}$$

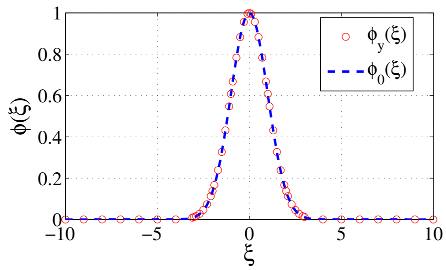

Fig. 2. $\phi_{\mathbf{y}_\infty}(\xi)$ versus $\phi_o(\xi)$.

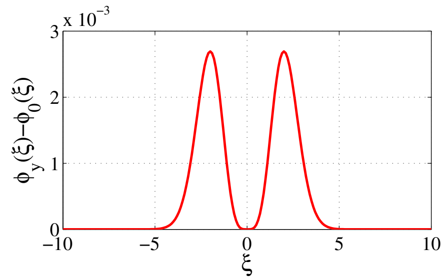

Fig. 3. $\phi_{\mathbf{y}_\infty}(\xi) - \phi_o(\xi)$ versus ξ.

where $\text{col}(\cdot)$ is a column vector obtained by stacking its arguments on top of each other and $\text{diag}(\cdot)$ is a (block) diagonal matrix obtained by placing its arguments on the diagonal. Suppose the regression data are independent of each other across the nodes, then

$$\boldsymbol{\mathcal{W}}_\infty \sim \mathbb{N}\left(\mathbf{1}_N \otimes w^o, \mathcal{P}\right) \tag{37}$$

where $\mathbf{1}_N$ is an all-one vector with N entries and $\otimes$ is the Kronecker product. We can express (37) in the following linear model form

$$\boldsymbol{\mathcal{W}}_\infty = \mathcal{H}w^o + \boldsymbol{\mathcal{V}} \tag{38}$$

where $\mathcal{H} \triangleq \mathbf{1}_N \otimes I_M$ and $\boldsymbol{\mathcal{V}}$ is zero mean (it actually consists of the weight error vectors across the nodes):

$$\boldsymbol{\mathcal{V}} \sim \mathbb{N}(0, \mathcal{P}) \tag{39}$$

Then, according to the Gauss-Markov Theorem [1], the MVUE of w^o is given by

$$\begin{aligned} \boldsymbol{w}^c &\triangleq (\mathcal{H}^T\mathcal{P}^{-1}\mathcal{H})^{-1}\mathcal{H}^T\mathcal{P}^{-1}\boldsymbol{\mathcal{W}}_\infty \\ &= \left(\sum_{k=1}^N P_k^{-1}\right)^{-1}\left(\sum_{k=1}^N P_k^{-1}\boldsymbol{w}_{k,\infty}\right) \end{aligned} \tag{40}$$

If the regressors across the network have uniform covariance matrix and step-size, i.e., $R_{u,k} = R_u$ and $\mu_k = \mu$, then (34) reduces to

$$\boldsymbol{w}_{k,\infty} \sim \mathbb{N}\left(w^o, \sigma_{v,k}^2 P_o\right) \tag{41}$$

where $P_o = \mu(2I - \mu R_u)^{-1}$, and (36) reduces to

$$\mathcal{P} = R_v \otimes P_o \tag{42}$$

where $R_v \triangleq \text{diag}(\sigma_{v,1}^2, \ldots, \sigma_{v,N}^2)$. Thus, (40) becomes

$$\begin{aligned} \boldsymbol{w}^c &= [(\mathbf{1}_N^T R_v^{-1}\mathbf{1}_N)^{-1} \otimes P_o][(\mathbf{1}_N^T R_v^{-1}) \otimes P_o^{-1}]\boldsymbol{\mathcal{W}}_\infty \\ &= \sum_{k=1}^N \frac{\sigma_{v,k}^{-2}}{\sum_{l=1}^N \sigma_{v,l}^{-2}}\boldsymbol{w}_{k,\infty} \end{aligned} \tag{43}$$

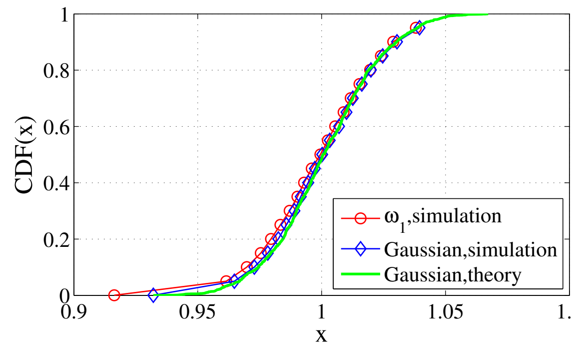

Fig. 4. Comparison of CDF between $[\boldsymbol{w}_\infty]_1$ and the corresponding Gaussian random variable.

The combination coefficients in (43) are proportional to the normalized signal-to-noise ratios across the nodes; this is of the same form as the MRC rule [1]. We conclude that MRC is a near-optimal fusion method.

5. SIMULATION RESULTS

We plot the characteristic function $\phi_{\mathbf{y}_\infty}(\xi)$ versus $\phi_o(\xi)$ in Fig. 2, and the difference $\phi_{\mathbf{y}_\infty}(\xi) - \phi_o(\xi)$ in Fig. 3 for $\rho = 0.995$. We see that the difference between $\phi_{\mathbf{y}_\infty}(\xi)$ and $\phi_o(\xi)$ is negligible with respect to $\phi_{\mathbf{y}_\infty}(\xi)$. In Fig. 4, we plot the cumulative distribution function (CDF) of $[\boldsymbol{w}_\infty]_1$ for $M = 10$, $\mu = 0.01$, $\sigma_v^2 = 0.1$, $R_u = I$, and $[w^o]_1 = 1$. The CDF is obtained over 2000 trials. In each trial we take w_{1000} as the steady-state estimate. We also plot the simulated and theoretical CDF of a Gaussian random variable with mean $[w^o]_1 = 1$ and variance $\mu\sigma_v^2/(2-\mu) = 5.0 \times 10^{-4}$ where the simulated CDF is obtained over 2000 samples. We see that the CDF of $[\boldsymbol{w}_\infty]_1$ matches well that of the corresponding Gaussian random variable.

6. REFERENCES

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

[2] X. Zhao and A. H. Sayed, "Performance limits of LMS-based adaptive networks," in *Proc. ICASSP*, Prague, Czech Republic, May 2011.

[3] C. G. Lopes and A. H. Sayed, "Diffusion least-mean squares over adaptive networks: Formulation and performance analysis," *IEEE Trans. Signal Process.*, vol. 56, pp. 3122–3136, July 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] R. Bitmead, "Convergence in distribution of LMS-type adaptive parameter estimates," *IEEE Trans. Autom. Control*, vol. 28, pp. 54–60, Jan. 1983.

[6] H. J. Kushner and G. G. Yin, *Stochastic Approximation Algorithms and Applications*, Springer, NY, 1997.

[7] A. Papoulis and S. U. Pillai, *Probability, Random Variables and Stochastic Processes, 4th ed.*, McGraw-Hill, NY, 2002.

[8] I. S. Gradshteyn and I. M. Ryzhik, *Table of Integrals, Series, and Products (Seventh Edition)*, Elsevier, UK, 2007.