Steady-state performance of incremental learning over distributed networks for non-Gaussian data.

This item was submitted to Loughborough University's Institutional Repository by the/an author.

Citation: LI, L. ... et al., 2008. Steady-state performance of incremental learning over distributed networks for non-Gaussian data. IN: Proceedings of 2008 9th International Conference on Signal Processing (ICSP 2008), Beijing, China, 26-29 October, pp. 227-230.

Additional Information:

• This is a conference paper [© IEEE]. It is also available at: http://ieeexplore.ieee.org/ Personal use of this material is permitted. However, permission to reprint/republish this material for advertising or promotional purposes or for creating new collective works for resale or redistribution to servers or lists, or to reuse any copyrighted component of this work in other works must be obtained from the IEEE.

Metadata Record: https://dspace.lboro.ac.uk/2134/5586

Version: Published

Publisher: © IEEE

Please cite the published version.
This item was submitted to Loughborough’s Institutional Repository (https://dspace.lboro.ac.uk/) by the author and is made available under the following Creative Commons Licence conditions.

For the full text of this licence, please go to: http://creativecommons.org/licenses/by-nc-nd/2.5/
ICSP2008 Proceedings

Steady-State Performance of Incremental Learning over Distributed Networks for Non-Gaussian Data

Leilei Li* Yonggang Zhang* Jonathon A. Chambers* Ali H. Sayed*

*Department of Electronic and Electrical Engineering, Loughborough University
Loughborough LE11 3TU, UK, E-mail: L.Li2@lboro.ac.uk
+ Department of Electronic Engineering, University of California
Los Angeles, USA, CA 90095-1594

Abstract: In this paper, the steady-state performance of the distributed least mean-squares (dLMS) algorithm within an incremental network is evaluated without the restriction of Gaussian distributed inputs. Computer simulations are presented to verify the derived performance expressions.

Key words: Adaptive filters, distributed estimation, energy conservation, incremental algorithm

I. INTRODUCTION

A traditional centralized solution requires a powerful central processor and large amount of communications between nodes within a network. In order to reduce significantly the amount of processing and communications, distributed solutions, only exploiting local data exchanges and communications between immediate neighboring nodes, are desirable. The applications of such distributed adaptive networks would range from sensors to environmental monitoring, military surveillance, transportation and factory instrumentation [1, 2]. Reference [3] developed the dLMS algorithm and studied its performance under Gaussian data.

In this paper, we evaluate the steady-state performance of dLMS for non-Gaussian data using weighted spatial-temporal energy conservation arguments [3]-[6]. In particular, we derive theoretical expressions for the mean-square deviation (MSD), the excess mean-squared-error (EMSE) and the mean-square deviation (MSD), its performance under Gaussian data.

II. NETWORK MODEL AND THE dLMS ALGORITHM

As formulated in [3], an N-node network is considered, where each node has a separate zero-mean random complex valued desired response \( \mathbf{d}_k \) and a \( 1 \times M \) spatially distinct row input vector \( \mathbf{u}_k \) with zero-mean and random elements. Both \( \mathbf{d}_k \) and \( \mathbf{u}_k \) are assumed to be jointly wide-sense stationary, and \( k \) is used to denote the node index. At each node time realizations \( \{ \mathbf{d}_i, \mathbf{u}_i \} \) of \( \{ \mathbf{d}_i, \mathbf{u}_i \} \) are observed, where \( i \) indicates the discrete time index. We formulate the linear least-mean-squares estimation problem:

\[
\min_w J(w) \quad \text{and} \quad J(w) = E[\mathbf{d}^\top - \mathbf{U} \mathbf{W}]^2
\]

where the \( N \times 1 \) global desired response vector and \( N \times M \) input matrix are respectively:

\[
\mathbf{d} = \text{col}[\mathbf{d}_1, \mathbf{d}_2, \ldots, \mathbf{d}_N]
\]

\[
\mathbf{U} = \text{col}[\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_N].
\]

The solution \( \mathbf{w}^o \) satisfies the normal equations, [7]:

\[
\mathbf{R}_{dd} = \mathbf{R}_w \mathbf{w}^O
\]

where \( \mathbf{R}_{dd} = \mathbf{E} \mathbf{U}^\top \mathbf{d} \) and \( \mathbf{R}_w = \mathbf{E} \mathbf{U}^\top \mathbf{U} \).

Note that the cost function can be decomposed into, as [3]

\[
J(w) = \sum_{i=1}^{N} J_i(w) \quad \text{and} \quad J_i(w) = \| \mathbf{d}_i - \mathbf{u}_i \mathbf{w} \|^2
\]

Using this property, references [3, 4] proposed an incremental dLMS strategy with a cyclic estimation structure, as follows:

For each time instant \( i \geq 0 \):

\[
\begin{align*}
\mathbf{y}_i^{(l)} &= \mathbf{w}_{i-1}^{(l)} \\
\mathbf{y}_i^{(l)} &= \mathbf{y}_i^{(l)} + \mu_i \mathbf{u}_i \{ \mathbf{d}_i(\mathbf{w}_i^{(l)}) - \mathbf{u}_i \mathbf{y}_i^{(l)} \} \\
\mathbf{w}_i &= \mathbf{y}_i^{(l)}
\end{align*}
\]

where \( \mathbf{y}_i^{(l)} \) indicates a local estimate of \( \mathbf{w}^o \) at node \( k \) and time \( i \), and \( \mathbf{w}_i \) indicates the estimate at the same instant. For each time \( i \), each node utilizes the local data \( \{ \mathbf{d}_i, \mathbf{u}_i \} \) and \( \mathbf{y}_k^{(l)} \) received from node \( k-1 \) in the ring to obtain \( \mathbf{y}_i^{(l)} \). At the end of this cycle, \( \mathbf{y}_k^{(l)} \) is employed as both the global estimate \( \mathbf{w}_i \) and the initial condition for the next time instant.

III. PERFORMANCE ANALYSIS

To evaluate the performance of dLMS, the following assumptions are utilized:

A1) The relationship between the unknown system vector \( \mathbf{w}^o \) and \( \{ \mathbf{d}_i(\mathbf{w}_i^{(l)}) \} \) is:
\[ d_i(i) = u_i, w^s + v_i(i) \]  

where \( v_i(i) \) is a temporally and spatially white noise sequence with variance \( \sigma^2_{vi} \) independent of \( u_{ij} \) and \( d_j(j) \) for all \( i, j \).

A2) \( u_{ij} \) has spatial and temporal independence, namely \( u_{ik} \) is independent of \( u_{ij} \) for \( k \neq i \), and \( u_{ij} \) for \( i \neq j \).

Introduce the following local error signals defined as in [3] to carry out the evaluation:

\[ \bar{\Psi}^{(i)} = w^s - \Psi^{(i)}, \quad \bar{\Psi}^{(i)} = w^s - \Psi^{(i)} \]

\[ e_{ik}(i) = u_{ik} \bar{\Psi}^{(i)}, \quad e_{kk}(i) = u_{ik} \bar{\Psi}^{(i)} \]

Note that the output error is given by

\[ e_i(i) = e_{ik}(i) + v_i(i) \]

We are interested in evaluating the following performance measures at each node \( k \):

\[ \eta_k = E[\bar{\Psi}^{(i)}] \quad \text{(MSD)} \]

\[ \xi_k = E[u_i, \bar{\Psi}^{(i)}] = E[e_{ik}(i)] \quad \text{(EMSE)} \]

\[ \xi_k = \xi_k + \sigma^2_{\delta k} \quad \text{(MSE)}. \]

Introduce further weighted error signals:

\[ e_{ik}^{(i)}(i) = u_{ik} \Sigma_k \bar{\Psi}^{(i)}, \quad e_{kk}^{(i)}(i) = u_{ik} \Sigma_k \bar{\Psi}^{(i)} \]

where \( \Sigma_k \) is a Hermitian positive-definite weighting matrix that we are free to choose at each node \( k \). Introduce also the weighted norm notation \( \|x\|_o = x^* \Sigma x \) for a vector \( x \) and Hermitian positive-definite \( \Sigma > 0 \). After the same manipulations as in [3], we obtain

\[ \bar{\Psi}^{(i)} + u_{ik} \bar{e}_{ik}(i) = \bar{\Psi}^{(i)} + u_{ik} \bar{e}_{ik}(i) \]

which reveals a relationship between two neighbouring nodes. By calculating the energies of both sides of (14), a spatial-temporal energy relation is obtained as

\[ E[\bar{\Psi}^{(i)}] = E[\bar{\Psi}^{(i)}] + \mu_e \sigma^2_{\delta k} E[\bar{\Psi}^{(i)}] \]

(16)

where \( \Sigma_k \) is given by

\[ \Sigma_k = \Sigma_k + \mu_e E[u_i u_i] \]

(17)

In order to evaluate the performance of the learning algorithm, we need to examine the following three moments:

\[ E[u_i u_i] = R_{kk}, E[u_i, u_i] = R_{kk}, E[u_i, u_i] \]

(18)

where \( \text{Tr}(A) \) denote the trace of a matrix \( A \).

In [3, 4], the assumption of Gaussian regressors is used to evaluate the last item in (18). In the non-Gaussian case, we proceed as in [5, 6] by introducing the \( M \times 1 \) vectors:

\[ \sigma_k = \text{vec} \Sigma_k \quad \text{and} \quad \delta_k = \text{vec} \Sigma_k \]

(19)

where we use the vec notation in two ways: \( \delta = \text{vec} \Sigma \) denotes an \( M \times 1 \) column vector whose entries are formed by stacking the successive columns of an \( M \times M \) matrix on top of each other, and \( \Sigma = \text{vec} \Sigma \) indicates a matrix whose entries are recovered from \( \delta \). We exploit the following useful property for the vec notation when working with Kronecker products: for any matrices \( \{P, \Sigma, \Omega\} \) of compatible dimensions, it holds that

\[ \text{vec}(P \Sigma \Omega) = (P^\top \otimes \Omega) \text{vec}(\Sigma). \]

(20)

Using (20) to express some items in (17), we find

\[ \text{vec}(E[u_i u_i] \Sigma_k) = (I \otimes R_{kk}) \delta_k \]

(21)

\[ \text{vec}(\Sigma_k, E[u_i u_i]) = (R_{kk}^\top \otimes I) \delta_k \]

(22)

\[ \text{vec}(E[u_i u_i]) = \text{vec}(u_i u_i) \otimes (u_i u_i) \delta_k. \]

(23)

Applying the vec notation to both sides of (17), a linear relation between the corresponding vectors \( \{\delta_k, \delta_k\} \) is obtained, namely

\[ \delta_k = F_k \delta_k \]

(24)

where \( F_k \) is an \( M^2 \times M^2 \) matrix and given by

\[ F_k = I - \mu_e (R_{kk}^\top \otimes I) - \mu_e (I \otimes R_{kk}) + \mu_e E[u_i u_i] \]

(25)

For clarity, we recall the time index \( i \). Therefore, expression (16) becomes

\[ E[\bar{\Psi}^{(i)}] = E[\bar{\Psi}^{(i)}] + \mu_e \sigma^2_{\delta k} \delta_k \]

(26)

where we use the fact that \( E[u_i u_i] \delta_k = \text{Tr}(R_{kk} \Sigma_k) = \delta_k \delta_k \) with \( \delta_k = \text{vec}(R_{kk}) \). For simplicity of notation, we drop the vec notation.
notation from the subscripts in (26):

\[
E[\nu_k^2] = E[\nu_k^2] + \mu_k^2 \sigma_k^2 (p_k^T \delta_k)
\] (27)

Let \( \mathbf{p}_k = \nu_k^2 \). Then

\[
E[\mathbf{p}_k] = E[\mathbf{p}_k] + \mu_k^2 \sigma_k^2 (p_k^T \delta_k)
\] (28)

By iterating (28) over one cycle, N coupled equations are obtained:

\[
E[\mathbf{p}_k] = E[\mathbf{p}_k] + g_k \delta_k
\] (29)

\[
E[\mathbf{p}_{k-1}] = E[\mathbf{p}_{k-1}] + g_{k-1} \delta_{k-1}
\] (30)

\[
E[\mathbf{p}_k] = E[\mathbf{p}_k] + g_k \delta_k
\] (31)

\[
E[\mathbf{p}_{k-1}] = E[\mathbf{p}_{k-1}] + g_{k-1} \delta_{k-1}
\] (32)

with \( g_k = \mu_k^2 \sigma_k^2 (p_k^T \delta_k) \). By choosing the free parameters \( \delta_{k-1} \), such that \( \delta_{k-1} = F_k \delta_k \), we combine (30) and (31) to obtain

\[
E[\mathbf{p}_k] = E[\mathbf{p}_k] + g_k \delta_k
\] (33)

Iterating across the cycle we arrive at

\[
E[\mathbf{p}_{k-1}] = E[\mathbf{p}_{k-1}] + g_{k-1} \delta_{k-1}
\] (34)

Let

\[
\Pi_{k-1,l} = F_{k-1,l} F_{k-1,l} \ldots F_{k-1,l} \delta_{k-1} \quad \text{for } l = 1, \ldots N
\] (35)

\[
a_{k-1} = g_k \Pi_{k-1,1} + g_{k-1} \Pi_{k-1,2} + \ldots + g_{k-2} \Pi_{k-1,N} + g_{k-1} \delta_{k-1}
\] (36)

Then

\[
E[\mathbf{p}_{k-1}] = a_{k-1} \delta_{k-1}
\] (37)

Expression (37) can be exploited to evaluate the performance measures at node k, as follows:

\[
\eta_k = E[\mathbf{p}_{k-1}] \quad q = \text{vec} \{ \mathbf{r}_k \} \quad \text{(MSD)}
\] (38)

\[
\xi_k = E[\mathbf{p}_{k-1}] \quad \text{(EMSE)}
\] (39)

\[
\xi_k = E[\mathbf{p}_{k-1}] + \sigma_{k-1}^2 \quad \text{(MSE)}
\] (40)

Since we are free to select the weight vector \( \delta_{k-1} \), choosing \( \delta_{k-1} = (I - \Pi_{k-1,1})^T q \) or \( \delta_{k-1} = (I - \Pi_{k-1,1})^T r_k \) results in expressions for the steady-state MSD, EMSE and MSE at node k

\[
\eta_k = a_{k-1} (I - \Pi_{k-1,1})^T q \quad \text{(MSD)}
\] (42)

\[
\xi_k = a_{k-1} (I - \Pi_{k-1,1})^T r_k \quad \text{(EMSE)}
\] (42)

\[
\xi_k = a_{k-1} (I - \Pi_{k-1,1})^T r_k \quad \text{(MSE)}
\] (43)

Fig.1  Statistical profile of the Gaussian data input to a 20-node network

Fig.2  Statistical profile of the uniform data input to a 20-node network
IV. SIMULATIONS

In this section, we compare the theoretical performance with the computer simulations in a system identification scenario. All simulation results are averaged over 100 independent runs. The steady-state curves are obtained by averaging the last 2000 instantaneous samples of 20000 iterations. Consider a network with 20 nodes seeking an unknown filter with $M = 10$ taps. Two types of signals, Gaussian and uniformly distributed signals, are used to generated the inputs at each node $k$:

$$u_k(i) = \alpha_k u_k(i-1) + \beta_k \cdot \rho_k(i)$$  \hspace{1cm} (44)

which is a first-order autoregressive (AR) process with a pole at $\alpha_k$; $\rho_k(i)$ is a white, zero-mean, Gaussian random sequence with unit variance or a uniform random sequence between -1.0 and 1.0, $\alpha_k \in [0,1]$ and $\beta_k = \sqrt{\sigma_{\alpha_k}^2 (1 - \alpha_k^2)}$. In this way, the covariance matrix $R_{uk}$ of the regressor $u_k$ is a $10 \times 10$ Toeplitz matrix with entries $r_{ik}(m) = \sigma_{\alpha_k}^2 \alpha_k^{m-i}$, $m = 0, \ldots, M-1$ with $\sigma_{\alpha_k}^2 \in (0,1]$. The background noise is a white Gaussian process with variance $\sigma_{\beta_k}^2 \in [0,0.1]$. The node profiles for both Gaussian and uniformly distributed inputs are illustrated in Fig.1-2. Although there are some very small discrepancies between simulation and practice at certain nodes, Fig 3 shows that the theoretical results for MSE match well with the simulated results. In addition, such discrepancies can be reduced by decreasing the step-size, as can be seen in Fig4. Similar results for EMSE and MSD have been obtained but are not included due to space limitations.

VII. CONCLUSIONS

In this paper, we carried out a steady-state mean square performance evaluation of dLMS under the assumptions A1 and A2. Using weighted spatial-temporal energy conservation arguments, we derived expressions for the steady-state MSD, EMSE and MSE without restricting the distribution of the inputs.

ACKNOWLEDGMENT

The work of A. H. Sayed was supported in part by NSF grants ECS-0601266 and ECCS-0725441.

REFERENCES