Mean Weight Behavior of the NLMS Algorithm for Correlated Gaussian Inputs

Tareq Y. Al-Naffouri, *Member, IEEE,* Muhammad Moinuddin, and Muhammad S. Sohail

***Abstract*—This paper presents a novel approach for evaluating the mean behavior of the well known normalized least mean squares (NLMS) adaptive algorithm for a circularly correlated Gaussian input. The mean analysis of the NLMS algorithm requires the calculation of some normalized moments of the input. This is done by first expressing these moments in terms of ratios of quadratic forms of spherically symmetric random variables and finding the cumulative density function (CDF) of these variables. The CDF is then used to calculate the required moments. As a result, we obtain explicit expressions for the mean behavior of the NLMS algorithm.**

***Index Terms*—Adaptive algorithms, mean behavior, spherically symmetric random variables, indefinite quadratic forms.**

I. Introduction

It is well known that the NLMS algorithm demonstrates its value as compared to the LMS algorithm for correlated inputs. This has been mainly observed by simulation [1], [2]. While the performance of LMS is well understood for a correlated input [3], [4], the same cannot be said about the NLMS algorithm. Thus, several works have attempted to study the performance of NLMS for correlated Gaussian input. However, the corresponding analyses either do not result in closed-form performance expressions [5]-[7] or rely on strong assumptions. Examples of such assumptions include the separation principle [8], [9], approximations [10], white input [5], [10], [11], specific structure of input regressor's distribution [2], [9], [12], small step size [9], long filters [10] and approximate solutions using Abelian integrals [8], [13].

In this paper, we solve this problem in part by evaluating the mean performance of NLMS in closed form for circularly symmetric correlated Gaussian input. Mean analysis requires evaluating some normalized moments of the input. We evaluate these moments by rewriting each as a ratio of the weighted norms of spherically symmetric random variables and finding its CDF. This is done in turn by rewriting this variable in indefinite quadratic form and using complex integration to obtain the CDF. This CDF can then be used to evaluate the multidimensional moments involved in closed form. The mean and mean square analysis of the ϵ-NLMS algorithm was considered in [14]. Unfortunately the results obtained cannot be specialized to the NLMS case by setting $\epsilon = 0$, as that results in indefinite expansions[1] of the form $\infty - \infty$ which are difficult to resolve. As pointed out by one of the reviewers, this moment has been derived in [15] based on a technique first introduced by [16]. Nevertheless, the derivation we present here is still significant as [15] derives it in a different context. Moreover, the approach we pursue here is drastically different from that of [15], [16] and we claim, is much simpler. In addition, our approach can be extended to derive second-order moments that are needed for mean-square analysis of NLMS.

The main contributions of this paper are:

a) The analysis presented is generalized in that it is not restricted to a specific input correlation matrix and does not rely on any assumptions[2] such as small step-size, white input or independence between regressor elements.
b) Closed form transient analysis of the mean behavior of the NLMS algorithm is carried out in a transparent manner.
c) The analysis allows us to evaluate the optimum step size in closed form for faster convergence in the mean.

The paper is organized as follows. Section II presents the system model. Section III derives the moments needed to evaluate the mean behavior and derives the optimal step size in terms of these moments. Section IV shows how these moments can be derived from the CDF of spherically symmetric random variables. Simulation results investigating the performance and validating the analytical model are presented in Section V. Finally, Section VI presents concluding remarks.

II. System Model

The weight update equation of the NLMS algorithm is given as[3]

$$\mathbf{w}(i+1) = \mathbf{w}(i) + \mu \frac{\mathbf{u}^{\mathrm{H}}(i)}{||\mathbf{u}(i)||^2} e(i), \qquad i \geq 0 \tag{1}$$

where $\mathbf{w}(i)$ is an estimate of the desired weight column vector $\mathbf{w}^{\mathrm{o}}$ (starting from some $\mathbf{w}(0)$ usually set to zero), $\mathbf{u}(i)$ is the input regression row vector and $e(i)$ is the estimation error defined as

$$e(i) = d(i) - \mathbf{u}(i)\mathbf{w}(i) = \mathbf{u}(i)\mathbf{w}^{\mathrm{o}} - \mathbf{u}(i)\mathbf{w}(i) + v(i) \tag{2}$$

where $v(i)$ is a zero-mean i.i.d noise independent of the input sequence with variance σ_v^2. It is more convenient to express

T. Y. Al-Naffouri and M. S. Sohail are with the Electrical Engineering Department, King Fahd University of Petroleum & Minerals, Dhahran, Saudi Arabia (alnaffouri@kfupm.edu.sa; saqib@kfupm.edu.sa).

M. Moinuddin is with the Telecommunication Engineering Department, Iqra University, Karachi, Pakistan (mmoin73@yahoo.com).

[1] In fact, the expressions obtained in the NLMS case are completely different in form from those of ϵ-NLMS, indicating the need for a different approach for the $\epsilon = 0$ case.

[2] Our analysis is exact up to the independence assumption that is usually used in adaptive filters.

[3] $(.)^H$ denotes conjugate transpose.

the adaptation equation (1) in terms of the weight error vector $\tilde{\mathbf{w}}(i) = \mathbf{w}^{\text{o}} - \mathbf{w}(i)$ as

$$\tilde{\mathbf{w}}(i+1) = \tilde{\mathbf{w}}(i) - \mu \frac{\mathbf{u}^{\text{H}}(i)}{||\mathbf{u}(i)||^2} e(i), \qquad i \geq 0 \tag{3}$$

and the estimation error as

$$e(i) = \mathbf{u}(i)\tilde{\mathbf{w}}(i) + v(i). \tag{4}$$

The analysis assumes that the independence assumption is valid. We will restrict our attention to circularly symmetric Gaussian inputs, i.e. $\mathbf{u}(i) \sim \mathcal{CN}(\mathbf{0}, \mathbf{R})$. Here we do not assume any specific structure for $\mathbf{R}$. Now, consider the eigenvalue decomposition of $\mathbf{R} = \mathbf{Q}^{\text{H}}\mathbf{\Lambda}\mathbf{Q}$ and define the rotated variables $\bar{\mathbf{w}}(i) = \mathbf{Q}\tilde{\mathbf{w}}(i)$ and $\bar{\mathbf{u}}(i) = \mathbf{u}(i)\mathbf{Q}^{\text{H}}$. Then by premultiplying both the sides of (1) by $\mathbf{Q}$, we get

$$\bar{\mathbf{w}}(i+1) = \bar{\mathbf{w}}(i) - \mu \frac{\bar{\mathbf{u}}(i)^{\text{H}}}{||\bar{\mathbf{u}}(i)||^2} e(i), \qquad i \geq 0 \tag{5}$$

where in arriving at (5), we have used the fact that $||\mathbf{u}(i)||^2 = ||\bar{\mathbf{u}}(i)\mathbf{Q}^{\text{H}}||^2 = ||\bar{\mathbf{u}}(i)||^2$. Note that $e(i)$ in (5) can be expressed in terms of the rotated vectors as

$$e(i) = \mathbf{u}(i)\mathbf{Q}^{\text{H}}\mathbf{Q}\tilde{\mathbf{w}}(i) + v(i) = \bar{\mathbf{u}}(i)\bar{\mathbf{w}}(i) + v(i). \tag{6}$$

The adaptive filter is thus solely a function of $\bar{\mathbf{u}}(i)$, which has a diagonal correlation matrix $\mathbf{\Lambda}$.

III. Mean Behavior of the NLMS Algorithm

Substituting the value of estimation error from (6), the recursion for the transformed mean weight error vector of the NLMS algorithm, given in (5), can be rewritten as

$$\bar{\mathbf{w}}(i+1) = \left[\mathbf{I} - \mu \left(\frac{\bar{\mathbf{u}}^{\text{H}}(i)\bar{\mathbf{u}}(i)}{||\bar{\mathbf{u}}(i)||^2}\right)\right] \bar{\mathbf{w}}(i) - \mu \frac{\bar{\mathbf{u}}^{\text{H}}(i)v(i)}{||\bar{\mathbf{u}}(i)||^2}. \tag{7}$$

Taking the expectation of both sides and assuming that the sequence $\{u(i)\}$ is i.i.d[4], it is easy to see that the mean behavior of the weight error vector is governed by the following recursion

$$E[\bar{\mathbf{w}}(i+1)] = [\mathbf{I} - \mu\mathbf{E}_1] E[\bar{\mathbf{w}}(i)] \tag{8}$$

with

$$\mathbf{E}_1 = E\left[\frac{\bar{\mathbf{u}}^{\text{H}}(i)\bar{\mathbf{u}}(i)}{||\bar{\mathbf{u}}(i)||^2}\right]. \tag{9}$$

We need to evaluate the matrix of moments $\mathbf{E_1}$. The off-diagonal entries, given by $E_{kk'} = E\left[\bar{u}_k(i)\bar{u}_{k'}(i)/||\bar{\mathbf{u}}(i)||^2\right]$ $(k = 1, 2, \cdots, M)$, are zero because $E_{kk'}$ is an odd function of $\bar{u}_k(i)$, which has a symmetric probability density function (pdf) and is independent of the rest of the elements of $\bar{\mathbf{u}}(i)$. Thus, the moment matrix $\mathbf{E_1}$ is a diagonal matrix. Let $\rho_k = E\left[|\bar{u}_k(i)|^2/||\bar{\mathbf{u}}(i)||^2\right]$ denote its kth diagonal value, then the kth entry of $\bar{\mathbf{w}}(i)$ is given by

$$E[\bar{w}_k(i)] = (1 - \mu\rho_k)^{i+1} E[\bar{w}_k(-1)], \qquad i \geq 0 \tag{10}$$

The term $(1 - \mu\rho_k)$ is referred to as the mode associated with $E[\bar{w}_k(i)]$. Thus, a necessary and sufficient condition for convergence is $|1 - \mu\rho_k| < 1,$ for all k, or equivalently

$$0 < \mu < \frac{2}{\rho_{\max}}. \tag{11}$$

[4]This is one possible form of the independence assumption.

We can also determine the step size that guarantees fastest convergence (by minimizing the largest mode [6])

$$\mu^{\text{o}} = \min_{\mu} \max_{k} |1 - \mu\rho_k| = \frac{2}{\rho_{\max} + \rho_{\min}}. \tag{12}$$

IV. Our Approach

It is clear that the mean performance of the NLMS algorithm is completely characterized by the moment matrix $\mathbf{E_1}$. Here onwards, we will drop the dependence of variables on the time index i for notational convenience. Now the entries of the moment matrix $\mathbf{E_1}$ can be determined from the expectation of the following random variable

$$s_k = \frac{|\bar{u}_k|^2}{||\bar{\mathbf{u}}||^2}, \qquad k = 1, 2, \cdots, M \tag{13}$$

The key to determining this expectation is to first define s_k in terms of spherically symmetric random variables. To do this, let $\hat{u}_k$ be the normalized version of $\bar{u}_k$, i.e. $\bar{u}_k = \sqrt{\lambda_k}\hat{u}_k$, where λ_k is the kth eigenvalue of the autocorrelation matrix of the input sequence, then we can rewrite s_k as

$$s_k = \frac{\lambda_k|\hat{u}_k|^2}{\sum_{j=1}^{M}\lambda_j|\hat{u}_j|^2}. \tag{14}$$

In Appendix A, we show that the CDF of s_k can be written as

$$F_{s_k}(x) = \sum_{m=1}^{M} \frac{(\lambda_m x - \sigma_{m,k})^{M-1} h(\lambda_m x - \sigma_{m,k})}{\Pi_{j=1,j\neq m}^{M}[(\sigma_{j,k} - \sigma_{m,k}) - (\lambda_j - \lambda_m)x]}. \tag{15}$$

The parameter $\sigma_{m,k}$ and function $h(.)$ are defined in Appendix A. We now use this CDF to determine the first moment of s_k. From (14), it is easy to see that s_k has support over the interval $(0, 1)$. Thus, the first moment of s_k can be expressed using integration by parts as

$$E[s_k] = \int_0^1 [1 - F_{s_k}(x)]\, dx = 1 -$$
$$\sum_{m=1}^{M} \int_{\frac{\sigma_{m,j}}{\lambda_m}}^{1} \frac{(\lambda_m x - \sigma_{m,k})^{M-1}}{\Pi_{j=1,j\neq m}^{M}[(\sigma_{j,k} - \sigma_{m,k}) - (\lambda_j - \lambda_m)x]} dx. \tag{16}$$

Applying partial fraction expansion to solve the above integral, we get

$$\frac{(\lambda_m x - \sigma_{m,k})^{M-1}}{\Pi_{j=1,j\neq m}^{M}[(\sigma_{j,k} - \sigma_{m,k}) - (\lambda_j - \lambda_m)x]} =$$
$$\frac{\lambda_m^{M-1}}{\Pi_{j=1,j\neq m}^{M}(\lambda_m - \lambda_j)} \left[1 + \sum_{j=1,j\neq m}^{M} \frac{c_j}{x + \bar{\sigma}_{jm}}\right] \tag{17}$$

where $\bar{\sigma}_{jm} = (\sigma_{j,k} - \sigma_{m,k})/(\lambda_m - \lambda_j)$ and

$$c_j = \frac{(-\bar{\sigma}_{jm} - \frac{\sigma_{m,k}}{\lambda_m})^{M-1}}{\Pi_{l=1,l\neq j,m}^{M}(\bar{\sigma}_{lm} - \bar{\sigma}_{jm})}. \tag{18}$$

Ultimately, the first moment of s_k is found to be

$$E[s_k] = 1 - \sum_{m=1}^{M} \frac{\lambda_m^{M-1}}{\Pi_{j=1,j\neq m}^{M}(\lambda_m - \lambda_j)} \Bigg[1 - \frac{\sigma_{m,k}}{\lambda_m} + \sum_{j=1,j\neq m}^{M} c_j \ln\left(\frac{1 + \bar{\sigma}_{jm}}{\frac{\sigma_{m,k}}{\lambda_m} + \bar{\sigma}_{jm}}\right)\Bigg]. \tag{19}$$

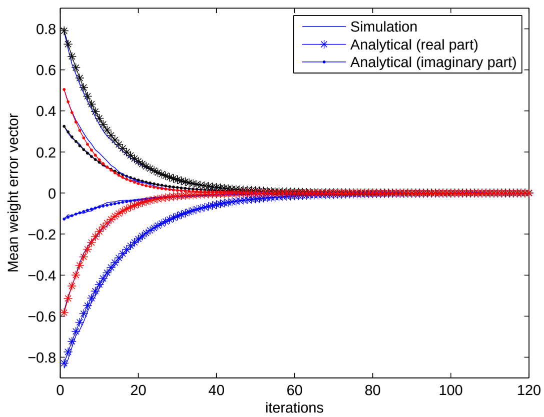

Fig. 1. Learning curve of the real and imaginary parts of first three taps of the mean weight error vector ($\mu_{\text{opt}} = 0.7627$).

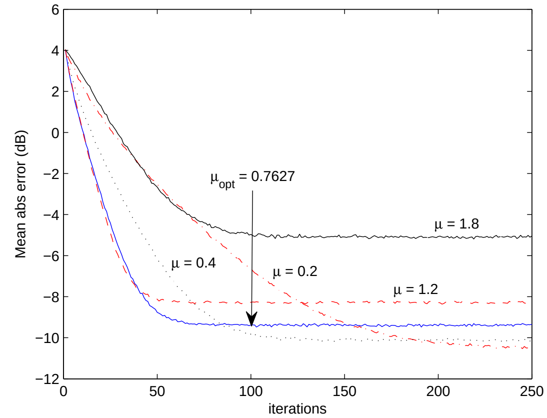

Fig. 2. Comparision of mean absolute error for various μ.

V. Simulation Results

The system noise is assumed to be a zero-mean i.i.d. sequence with variance 0.01. The length of both the unknown system and the adaptive filter used is assumed to be 5. The input to the adaptive filter and unknown system is correlated complex Gaussian noise with correlation matrix

$$\mathbf{R} = \begin{bmatrix} 1 & \alpha_c & \alpha_c^2 & \cdots & \alpha_c^{M-1} \\ \alpha_c & 1 & \alpha_c & \cdots & \alpha_c^{M-2} \\ \alpha_c^2 & \alpha_c & 1 & \cdots & \alpha_c^{M-3} \\ \vdots & & & & \\ \alpha_c^{M-1} & \alpha_c^{M-2} & \alpha_c^{M-3} & \cdots & 1 \end{bmatrix}$$

where $0 < \alpha_c < 1$ is the factor that controls the correlation between the regressor elements. All simulation results are obtained for $\alpha_c = 0.5$. The analytical result for the transient behavior of the mean weight error vector of the NLMS algorithm is investigated. Figure 1 depicts the analytical and simulation results for the learning curves of real and imaginary parts of the first three taps of the mean weight error vector, averaged over 200 runs. The results show excellent analytical tracking of the transient behavior of the taps, thereby validating the theoretical model proposed in the paper. Figure 2 shows the mean performance of the NLMS when employing the optimum step size, averaged over 5000 runs, demonstrating the value of choosing the step size optimally.

VI. Conclusion

We have presented a novel approach for mean analysis of the NLMS algorithm for correlated complex Gaussian input. Our approach reduces the mean analysis to determining the first multidimensional moment. Our approach shows that this moment can be determined from the CDF of a variable of the form $[||\phi||^2_{\mathbf{A}}]/[||\phi||^2_{\mathbf{B}}]$. The advantage of this approach is its transparency and its ability to evaluate performance in closed form. Our theoretical findings of the transient performance are corroborated by simulation results. This approach can also be extended to mean-square analysis of the NLMS algorithm.

Appendix

We define the random vector

$$\boldsymbol{\phi} = \left[\frac{\hat{u}_1}{\|\hat{\mathbf{u}}\|}, \frac{\hat{u}_2}{\|\hat{\mathbf{u}}\|}, \cdots, \frac{\hat{u}_M}{\|\hat{\mathbf{u}}\|}\right]^{\mathrm{T}} \tag{20}$$

and rewrite (14) as a ratio of weighted norms of $\boldsymbol{\phi}$, i.e.[5]

$$s_k = \frac{\|\boldsymbol{\phi}\|^2_{\boldsymbol{\Sigma}_k}}{\|\boldsymbol{\phi}\|^2_{\boldsymbol{\Lambda}}}, \tag{21}$$

where $\boldsymbol{\Sigma}_k$ is an $M \times M$ matrix with all elements equal to zero except the kth element in the main diagonal, which is set to λ_k. The random vector (20) is known as a spherically symmetric random vector [17] and has the pdf

$$p(\boldsymbol{\phi}) = \frac{M!}{\pi^M}\delta(\|\boldsymbol{\phi}\|^2 - 1) \tag{22}$$

where $\delta(.)$ is the delta function. We will now show how (21) in conjunction with (22) can be used to determine the distribution of s_k. The CDF of s_k, denoted by $F_{s_k}(x)$, is defined as

$$\begin{aligned} F_{s_k}(x) &= P(s_k \leq x) && (23) \\ &= P\left(\frac{\|\boldsymbol{\phi}\|^2_{\boldsymbol{\Sigma}_k}}{\|\boldsymbol{\phi}\|^2_{\boldsymbol{\Lambda}}} \leq x\right) && (24) \\ &= P(\|\boldsymbol{\phi}\|^2_{x\boldsymbol{\Lambda}-\boldsymbol{\Sigma}_\mathbf{k}} \geq 0) && (25) \end{aligned}$$

thus

$$F_{s_k}(x) = \int_{\|\boldsymbol{\phi}\|^2_{x\boldsymbol{\Lambda}-\boldsymbol{\Sigma}_k \geq 0}} p(\boldsymbol{\phi})d\boldsymbol{\phi}. \tag{26}$$

This is an M-dimensional integral over the region defined by the inequality $\|\boldsymbol{\phi}\|^2_{x\boldsymbol{\Lambda}-\boldsymbol{\Sigma}_k \geq 0}$ which is difficult to evaluate. We can write (26) as an unconstrained integral by using the unit step function as

$$F_{s_k}(x) = \int p(\boldsymbol{\phi})h(\|\boldsymbol{\phi}\|^2_{x\boldsymbol{\Lambda}-\boldsymbol{\Sigma}_k})d\boldsymbol{\phi} \tag{27}$$

where $h(.)$ is the step function. This integral is still difficult to solve due to the presence of the delta and step functions. To

[5]The weighted norm is defined by $\|\alpha\|^2_{\mathbf{A}} \triangleq \alpha^{\mathrm{H}}\mathbf{A}\alpha$.

get around this, we replace them with the following equivalent integral representation (see [17], [18] for details)

$$\delta(\|\boldsymbol{\phi}\|^2-1) = \frac{1}{2\pi}\int e^{(\alpha+j\omega)(\|\boldsymbol{\phi}\|^2-1)}d\omega \quad (28)$$

$$h(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty}\frac{e^{x(j\omega+\beta)}}{j\omega+\beta}d\omega \quad (29)$$

which are valid for any $\alpha, \beta > 0$, where α and β are free parameters that we can choose conveniently. After replacing delta and step functions with their equivalent integral representations, the CDF of s_k can be set up as

$$F_{s_k}(x) = \frac{M!}{4\pi^{M+2}}\ e^{\alpha}\int_{-\infty}^{\infty}\frac{1}{j\omega_1+\beta}d\omega_1\int_{-\infty}^{\infty}e^{-j\omega_2}d\omega_2 \int e^{-\boldsymbol{\phi}^{\mathrm{H}}[\alpha\mathbf{I}+(\boldsymbol{\Sigma}_k-x\boldsymbol{\Lambda})(j\omega_1+\beta)-j\omega_2\mathbf{I}]\boldsymbol{\phi}}d\boldsymbol{\phi}. \quad (30)$$

By inspecting the inner integral, we note that it is similar to the Gaussian density integral. Intuition suggests that (see [18] where we give a rigorous derivation for a similar integral)

$$\frac{1}{\pi^M}\int e^{-\boldsymbol{\phi}^{\mathrm{H}}[\alpha\mathbf{I}+(\boldsymbol{\Sigma}_k-x\boldsymbol{\Lambda})(j\omega_1+\beta)-j\omega_2\mathbf{I}]\boldsymbol{\phi}}d\boldsymbol{\phi} = \frac{1}{\left|\alpha\mathbf{I}+(\boldsymbol{\Sigma}_k-x\boldsymbol{\Lambda})(j\omega_1+\beta)-j\omega_2\mathbf{I}\right|}. \quad (31)$$

This leads to a 2-D integral in ω_1 and ω_2

$$F_{s_k}(x) = \frac{M!}{4\pi^2}\ e^{\alpha}\int_{-\infty}^{\infty}\frac{1}{j\omega_1+\beta}d\omega_1 \int_{-\infty}^{\infty}\frac{e^{-j\omega_2}}{\left|\alpha\mathbf{I}+(\boldsymbol{\Sigma}_k-x\boldsymbol{\Lambda})(j\omega_1+\beta)-j\omega_2\mathbf{I}\right|}d\omega_2. \quad (32)$$

To evaluate the above integral w.r.t. ω_2, we use partial fraction expansion to represent the determinant in (32) as

$$\frac{1}{(j\omega_1+\beta)^{M-1}}\sum_{m=1}^{M}\frac{\eta_m}{\alpha+(\sigma_{m,k}-\lambda_m x)(j\omega_1+\beta)-j\omega_2} \quad (33)$$

where $\sigma_{m,k}$ and λ_m represent the mth diagonal elements of the matrices $\boldsymbol{\Sigma}_k$ and $\boldsymbol{\Lambda}$, respectively, and the constant

$$\eta_m = \frac{1}{\Pi_{l=1,l\neq m}^{M}[(\sigma_{l,k}-\sigma_{m,k})-(\lambda_l-\lambda_m)x]}. \quad (34)$$

We can now use residue theory to determine the integral with respect to ω_2 as

$$\frac{1}{2\pi}\int_{-\infty}^{\infty}\frac{e^{-j\omega_2}}{\left|\alpha\mathbf{I}+(\boldsymbol{\Sigma}_k-x\boldsymbol{\Lambda})(j\omega_1+\beta)-j\omega_2\mathbf{I}\right|}d\omega_2 = e^{-\alpha}\sum_{m=1}^{M}\eta_m e^{-\sigma_{m,k}(x)(j\omega_1+\beta)}h[\alpha+\sigma_{m,k}(x)] \quad (35)$$

where

$$\sigma_{m,k}(x) = (\sigma_{m,k}-\lambda_m x). \quad (36)$$

The CDF of s_k can then be set up as

$$F_{s_k}(x) = \frac{M!}{2\pi}\int_{-\infty}^{\infty}\frac{1}{(j\omega_1+\beta)^M} \sum_{m=1}^{M}\eta_m e^{-\sigma_{m,k}(x)(j\omega_1+\beta)}h[\alpha+\sigma_{m,k}(x)]d\omega_1. \quad (37)$$

The above integral is then integrated with respect to ω_1, again using residue theory. The detailed algebraic manipulations are omitted due to lack of space. Finally, the CDF of s_k is given by

$$F_{s_k}(x) = \sum_{m=1}^{M}\frac{(\lambda_m x-\sigma_{m,k})^{M-1}h(\lambda_m x-\sigma_{m,k})}{\Pi_{j=1,j\neq m}^{M}[(\sigma_{j,k}-\sigma_{m,k})-(\lambda_j-\lambda_m)x]}. \quad (38)$$

Acknowledgement

This work was supported by King Abdulaziz City for Science & Technology (KACST), Saudi Arabia, Project no. 30-135.

References

[1] D. T. M. Slock, "On the convergence behavior of the LMS and the normalized LMS algorithms," *IEEE Trans. on Signal Process.*, vol. 41, no. 9, pp. 2811-2825, Sept. 1993.

[2] P. E. An, B. Brown, and C. J. Harris, "On the convergence rate performance of the normalized least-mean-square adaptation," *IEEE Trans. on Neural Networks*, vol. 8, no. 5, pp. 1211-1214, Sept. 1997.

[3] W. A. Gardner, "Learning characteristics of stochastic gradient descent algorithms: A general study, analysis and critique," *Signal Process.* vol. 6, no. 2, pp. 113-133, April 1984.

[4] A. Feurer and E. Weinstein "Convergence analysis of LMS filters with uncorrelated Gaussian data," *IEEE Trans. on Acoust., Speech, Signal Process.,* vol. ASSP-33, no. 1, pp. 222-230, Feb. 1985.

[5] S. C. Douglas and T. H. Y. Meng, "Normalized data nonlinearities for LMS adaptation," *IEEE Trans. on Signal Process.*, vol. 42, no. 6, pp. 1352-1365, June 1994.

[6] A. H. Sayed, *Fundamentals of Adaptive Filtering*. New York: Wiley-Interscience, 2003.

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

[8] E. M. Lobato, O. J. Tobias, and R. Seara, "Stochastic modeling of the transform-domain ϵ LMS algorithm," *IEEE Trans. on Signal Process.*, vol. 56, no. 5, pp. 1840-1852, May 2008.

[9] G. Barrault, M. H. Costa, J. C. M. Bermudez, and A. Lenzi, "A new analytical model for the NLMS algorithm," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.* (ICASSP), Philadelphia, PA, Mar. 2005, vol. 4, pp. 41-44.

[10] M. H. Costa and J. C. M. Bermudez, "An improved model for the normalized LMS algorithm with Gaussian inputs and large number of coefficients," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.* (ICASSP), Orlando, FL, May 2002, vol. II, pp. 1385-1388.

[11] V. H. Nascimento, "A simplified model for the effect of normalization on the convergance rate of adaptive filters," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.* (ICASSP), Montreal, Canada, May 2004, vol. II, pp. 453-456.

[12] M. Tarrab, and A. Feuer, "Convergence and performance analysis of the normalized LMS algorithm with uncorrelated Gaussian data," *IEEE Trans. on Inf. Theory*, vol. 34, no. 4, pp. 680-691, July 1988.

[13] J. E. Kolodziej, O. J. Tobias, and R. Seara, "An improved stochastic model of the nlms algorithm for correlated input data," in *Proc. European Signal Processing Conf.* (EUSIPCO), Pozna, Poland, Sept. 2007, pp. 360-364.

[14] T. Y. Al-Naffouri and M. Moinuddin, "Exact performance analysis of the ϵ-NLMS algorithm for colored circular Gaussian inputs," *IEEE Trans. on Signal Process.,* vol. 58, no. 10, pp. 5080-5090, Oct. 2010.

[15] J. E. Kolodziej, O. J. Tobias, R. Seara, and D. R. Morgan, "On the constrained stochastic gradient algorithm: model, performance, and improved version," *IEEE Trans. Signal Process.*, vol. 57, no, 4, pp. 1304-1315, Apr. 2009.

[16] M. Rupp, "The behaviour of LMS and NLMS algorithms in the presence of spherically invariant processes," *IEEE Trans. Signal Process.*, vol. 41, no. 3, pp. 1149-1160, Mar. 1993.

[17] B. Hassibi and T. L. Marzetta, "Multiple-antennas and isotropically-random unitary inputs: The received signal density in closed-form," *IEEE Trans. on Inf. Theory*, vol. 48, no. 6, pp. 1473-1484, June 2002.

[18] T. Y. Al-Naffouri and B. Hassibi, "On the distribution of indefinite quadratic forms in Gaussian random variables," in *Proc. Int. Symp. on Information Theory* (ISIT), South Korea, June 2009, pp. 1744-1748.