THE STEADY-STATE OF THE (NORMALIZED) LMS IS SCHUR CONVEX

Khaled A. Al-Hujaili[1] *Tareq Y. Al-Naffouri*[2] *Muhammad Moinuddin*[3,4]

[1] Department of Electrical Engineering, Taibah University,
Kingdom of Saudi Arabia, Email:khujaili@taibah.edu.sa
[2] Electrical Engineering Department, King Abdullah University of Science and Technology,
Kingdom of Saudi Arabia, Email:tareq.alnaffouri@kaust.edu.sa
[3]Electrical and Computer Engineering Department, King Abdulaziz University,
Kingdom of Saudi Arabia, Email: mmsansari@kau.edu.sa
[4]Center of Excellence in Intelligent Engineering Systems (CEIES), King Abdulaziz University,
Saudi Arabia, Email: mmsansari@kau.edu.sa

ABSTRACT

In this work, we demonstrate how the theory of majorization and schur-convexity can be used to assess the impact of input-spread on the Mean Squares Error (MSE) performance of adaptive filters. First, we show that the concept of majorization can be utilized to measure the spread in input-regressors and subsequently order the input-regressors according to their spread. Second, we prove that the MSE of the Least Mean Squares Error (LMS) and Normalized LMS (NLMS) algorithms are schur-convex, that is, the MSE of the LMS and the NLMS algorithms preserve the majorization order of the inputs which provide an analytical justification to why and how much the MSE performance of the LMS and the NLMS algorithms deteriorate as the spread in input increases.

***Index Terms*—** Adaptive Filters, Mean Square Error Analysis, Majorization, Schur-convexity, input-spread

1. INTRODUCTION

Both the least mean squares (LMS) and the Normalized LMS (NLMS) algorithms belong to the steepest-descent algorithm which is one of the most widely used adaptive algorithms due to its simplicity and robustness [1]. The LMS algorithm develops from an instantaneous approximation of the steepest-descent algorithm with the following weight update rule

$$\boldsymbol{w}_i = \boldsymbol{w}_{i-1} + \mu \boldsymbol{u}_i e_i, \;\; \boldsymbol{w}_{-1} = \text{initial guess}, \qquad (1)$$

where $\boldsymbol{w}_i$ and $\boldsymbol{u}_i$ represent $M \times 1$ filter-weights and input-regressor respectively, at time i. The regressor is assumed to be zero-mean complex-valued circular Gaussian with positive-definite covariance matrix $\boldsymbol{R}_u = \mathsf{E}\boldsymbol{u}\boldsymbol{u}^{\mathsf{H}}$, where $\mathsf{E}[\cdot]$ represents expectation operator. Further, $d_i = \boldsymbol{u}_i^{\mathsf{H}}\boldsymbol{w}^o + v_i$ is the reference signal at time i, $e_i = d_i - \boldsymbol{u}_i^{\mathsf{H}}\boldsymbol{w}_{i-1}$ is the estimation error, μ is *step-size*, v_i is white Gaussian noise with variance σ_v^2 and $\boldsymbol{w}^o$ is some arbitrary $M \times 1$ weight vector that we wish to estimate.

The NLMS algorithm uses the energy of the input-regressor to normalize the correction term in the weight update which results in the following weight update

$$\boldsymbol{w}_i = \boldsymbol{w}_{i-1} + \mu \frac{\boldsymbol{u}_i}{||\boldsymbol{u}_i||^2} e_i, \;\; \boldsymbol{w}_{-1} = \text{initial guess}, \qquad (2)$$

In order to assess the performance of adaptive algorithms, the most commonly used criterion is *steady-state MSE or excess MSE (EMSE)*. The steady-state EMSE is defined as EMSE $\triangleq \lim_{i\to\infty} \mathsf{E}|e_i|^2 - \sigma_v^2$. By employing energy-conservation relation with the aid of some assumptions on the data$\{d_i, \boldsymbol{u}_i\}$[1], the steady-state EMSE of the LMS (denoted by ζ_{LMS}) is [1]

$$\zeta_{LMS} = \frac{\sigma_v^2 \mu \sum_{k=1}^{M} \frac{\lambda_k}{2-\mu\lambda_k}}{1 - \mu \sum_{k=1}^{M} \frac{\lambda_k}{2-\mu\lambda_k}}, \qquad (3)$$

and an approximate[2] steady-state EMSE of the NLMS (denoted by ζ_{NLMS}) is [1]

$$\zeta_{NLMS} = \frac{\mu\sigma_v^2}{2-\mu}\left(\sum_{k=1}^{M}\lambda_k\right) E\left[\frac{1}{\|\boldsymbol{u}_i\|^2}\right], \qquad (4)$$

respectively. Here, $\{\lambda_k\}$ are the eigenvalues of $\boldsymbol{R}_u$ (sorted in descending order), i.e., $\lambda_m \geq \lambda_n, \forall\; 1 \leq m < n \leq M$. We collect $\{\lambda_k\}$ in an M-dimensional vector $\boldsymbol{\lambda}$ for notational convenience. Note that, from the EMSE point of view, the covariance matrix $\boldsymbol{R}_u$ can itself be assumed diagonal without losing any insight i.e., $\boldsymbol{R}_u = \boldsymbol{\Lambda} = \mathbf{d}iag(\boldsymbol{\lambda})$ [2].

[1]Due to space constraints, we refrain from including the procedure to obtain MSE and direct the interested reader to [1].

[2]Exact performance analysis of the NLMS can be found in [2, 3]. For the purpose of this paper, we confine our attention to the approximate expression in (4)

978-1-4799-9988-0/16/$31.00 ©2016 IEEE

2. QUANTIFICATION OF INPUT SPREAD AND MSE PERFORMANCE

In analyzing the behavior of adaptive filters, a common question is *"Given two inputs with correlation matrices $\mathbf{\Lambda}_1$ and $\mathbf{\Lambda}_2$[3], for which input does the adaptive algorithm perform better?""* Intuition suggests that the algorithm will perform better for the *less-spread* input. However, the notion that the input with covariance $\mathbf{\Lambda}_1$ is "less spread out" or "more nearly equal" than the input with covariance $\mathbf{\Lambda}_2$, is itself vague. Many researchers have attempted to study the performance of the LMS and the NLMS algorithms with the aim to understand why and how much the MSE performance of the two algorithms degrades to an increase in input correlation [2, 3, 4, 5]. In [2], closed form expressions for the transient analysis and the steady-state mean-square error (MSE) of the NLMS algorithm are developed but these expressions are in terms of multidimensional moments which [2] falls short of evaluating. Several other works have attempted to evaluate these moments but the corresponding analyses don't result in closed form performance expressions [1, 2, 4], or rely on strong assumptions [5, 6, 7, 8]. Towards this end, note that a commonly adopted measure to compare the spread of input signal is *eigenvalue-spread* [1], i.e., $\rho = \lambda_{\max}/\lambda_{\min}$. However, as evident from the definition of eigenvalue-spread, it is completely oblivious to any eigenvalue $\lambda_i \neq \{\lambda_{\max}, \lambda_{\min}\}$. Therefore, it is possible that two inputs have the same eigenvalue-spread (i.e., $\rho_1 = \rho_2$), and yet yield completely different EMSE, if $\boldsymbol{\lambda}_1 \neq \boldsymbol{\lambda}_2$ (numerical studies confirming this behaviour are included in *Simulation Results*).

In general, the notion that any vector $\boldsymbol{x}$ is more spread out than a vector $\boldsymbol{y}$, arises in a variety of contexts and can be made precise in a number of ways [9]. In remarkably many cases, the appropriate precise statement is that "$\boldsymbol{x}$ *majorizes* $\boldsymbol{y}$". We give a formal definition of this statement below:

Definition 1: For any two vectors $\boldsymbol{x} \in \mathbb{R}^M$ and $\boldsymbol{y} \in \mathbb{R}^M$ with descending order components $x_1 \geq x_2 \geq \cdots \geq x_M \geq 0$ and $y_1 \geq y_2 \geq \cdots \geq y_M \geq 0$, the vector $\boldsymbol{x}$ majorizes the vector $\boldsymbol{y}$ (written as $\boldsymbol{x} \succ \boldsymbol{y}$) if

$$\sum_{i=1}^{k} x_i \geq \sum_{i=1}^{k} y_i, \;\; k = 1, 2, \cdots, M-1, \; \textit{and} \; \sum_{i=1}^{M} x_i = \sum_{i=1}^{M} y_i. \tag{5}$$

In this work, we use majorization to compare the spread in input signals with eigenvalues collected in $\boldsymbol{\lambda}_1$ and $\boldsymbol{\lambda}_2$ respectively. Note that, we aim to study the impact of input-spread on the EMSE of the LMS and the NLMS algorithms, which is a *scalar-valued* function of $\boldsymbol{\lambda}$. Hence not only we are interested in *ordering* vectors based on their spread, but also in functions of these vectors that preserve this order (i.e., *order-preserving functions*). Specifically, the functions that preserve the ordering of majorization are known as *Schur-convex* functions. We give a formal definition of Schur-convex functions below:

Definition 2: A Schur-convex function, is a function $f : \mathbb{R}^M \to \mathbb{R}$, for which $\boldsymbol{x} \succ \boldsymbol{y}$ implies $f(\boldsymbol{x}) \geq f(\boldsymbol{y})$.

With the definition of majorization and Schur-convex functions, we are in a position to give the intuitive idea - that an increasing spread in input will deteriorate the EMSE - a mathematical form. Essentially, we need to show that if $\boldsymbol{\lambda}_1 \succ \boldsymbol{\lambda}_2$ (i.e., $\boldsymbol{\lambda}_1$ majorizes $\boldsymbol{\lambda}_2$), then $\zeta(\boldsymbol{\lambda}_1) \geq \zeta(\boldsymbol{\lambda}_2)$ and vice versa, or more simply that $\zeta(\boldsymbol{\lambda})$ is a Schur-convex function.

There are various theorems to investigate the Schur-Convexity of a function [9]. However, if the function $f(\boldsymbol{x})$ is symmetric (i.e., the elements of vector $\boldsymbol{x}$ can be arbitrarily permuted without changing the value of the function $f(\boldsymbol{x})$), then the test for Schur-convexity becomes simple and is outlined in the following lemma.

Lemma 1: The necessary and sufficient condition for a symmetric function $f(\boldsymbol{x})$ to be Schur-convex is

$$(x_1 - x_2)\left[\frac{\partial f(\boldsymbol{x})}{\partial x_1} - \frac{\partial f(\boldsymbol{x})}{\partial x_2}\right] \geq 0. \tag{6}$$

It is clear from the expressions of the EMSE given in (3) and (4)-that the EMSE of both the LMS and the NLMS are symmetric in $\boldsymbol{\lambda}$ and hence, we can use *Lemma 1* to check whether $\zeta_{LMS}(\boldsymbol{\lambda})$ and $\zeta_{NLMS}(\boldsymbol{\lambda})$ are Schur-convex or not, that is,

$$(\lambda_1 - \lambda_2)\left[\frac{\partial \zeta(\boldsymbol{\lambda})}{\partial \lambda_1} - \frac{\partial \zeta(\boldsymbol{\lambda})}{\partial \lambda_2}\right] \geq 0. \tag{7}$$

3. SCHUR-CONVEXITY OF EMSE OF THE LMS

In this section, we prove the Schur convexity of the LMS. To this end, rewrite (3) as

$$\zeta_{LMS}(\boldsymbol{\lambda}) = \frac{\sigma_v^2 \mu S(\boldsymbol{\lambda})}{(1 - \mu S(\boldsymbol{\lambda}))} \tag{8}$$

where $S(\boldsymbol{\lambda}) \triangleq \sum_{k=1}^{M} \frac{\lambda_k}{2 - \mu\lambda_k}$. Now, take the partial derivative of $\zeta_{LMS}(\boldsymbol{\lambda})$ w.r.t. λ_m to get

$$\frac{\partial \zeta_{LMS}(\boldsymbol{\lambda})}{\partial \lambda_m} = \frac{2\mu\sigma_v^2\Big[(1 - \mu S(\boldsymbol{\lambda}))\dfrac{\partial S(\boldsymbol{\lambda})}{\partial \lambda_m} - \mu S(\boldsymbol{\lambda})\dfrac{\partial S(\boldsymbol{\lambda})}{\partial \lambda_m}\Big]}{(1 - \mu S(\boldsymbol{\lambda}))^2}. \tag{9}$$

Since, we have

$$\frac{\partial S(\boldsymbol{\lambda})}{\partial \lambda_m} = \frac{\partial}{\partial \lambda_m}\left(\frac{\lambda_m}{2 - \mu\lambda_m} + \sum_{\substack{k=1\\k\neq m}}^{M} \frac{\lambda_k}{2 - \mu\lambda_k}\right) = \frac{2}{(2 - \mu\lambda_m)^2}. \tag{10}$$

[3] Here onwards we use $\mathbf{\Lambda}$ and $\boldsymbol{\lambda}$ interchangeably.

Using (9), (10) and *Lemma 1* (with appropriate replacements of x_1, x_2 and λ_m by λ_1 and λ_2), we get the following inequality to test the Schur-convexity of the $\zeta_{LMS}((\boldsymbol{\lambda}))$

$$\frac{2\mu\sigma_v^2}{(1-\mu S(\boldsymbol{\lambda}))^2}(\lambda_1-\lambda_2)\left[\frac{1}{(2-\mu\lambda_1)^2}-\frac{1}{(2-\mu\lambda_2)^2}\right]\overset{?}{\geq}0. \tag{11}$$

The first term - i.e., $2\mu\sigma_v^2(1-\mu S(\boldsymbol{\lambda}))^{-2}$ - is always positive and hence is insignificant in the test. Thus, the above test is reduced to

$$(\lambda_1-\lambda_2)\left[\frac{1}{(2-\mu\lambda_1)^2}-\frac{1}{(2-\mu\lambda_2)^2}\right]\overset{?}{\geq}0. \tag{12}$$

In the aforementioned inequality, if $\lambda_1 > \lambda_2$, then the first difference is positive i.e., $(\lambda_1-\lambda_2) > 0$. Similarly, $(2-s\mu\lambda_1)^2 < (2-\mu\lambda_2)^2 \Rightarrow 1/(2-\mu\lambda_1)^2 > 1/(2-\mu\lambda_2)^2$, and hence the second difference is also positive, guaranteeing that the inequality is satisfied. By similar reasoning, we note that if $\lambda_1 < \lambda_2$, both differences are negative, and hence their product (i.e., the left-hand-side of (12)) is again positive. Finally, the inequality is also satisfied if $\lambda_1 = \lambda_2$ and $(2-s\mu\lambda_p)^2 \neq 0$ where $p = 1, 2$. Note that, in the context of LMS adaptive filter, the stability condition requires that $\mu < \frac{2}{\lambda_{\max}}$ and hence $(2-\mu\lambda_p)^2 \neq 0$ is true for any $1 \leq p \leq M$.

4. SCHUR-CONVEXITY OF EMSE OF THE NLMS

We now turn our attention to the NLMS algorithm. Consider the EMSE expression in (4). To proceed, we need to express the moment $\Sigma_f \triangleq E\left[\frac{1}{\|\boldsymbol{u}_i\|^2}\right]$ in terms of Lambda. We do so by using the Taylor approximation with the first 3 terms only. This allows us to write

$$\Sigma_f = E\left[\frac{1}{\|\mathbf{u}\|^2}\right] \approx E\left[(\|\mathbf{u}\|^2)^2 - 3\|\mathbf{u}\|^2 + 3\right] \tag{13}$$

It can be shown that proving (7) is equivalent to prove

$$(\lambda_i-\lambda_j)\left(\frac{\partial\Sigma_f}{\partial\lambda_i}-\frac{\partial\Sigma_f}{\partial\lambda_j}\right)\geq 0 \tag{14}$$

Next, take the partial derivative of $\zeta_{NLMS}(\boldsymbol{\lambda})$ w.r.t. λ_m. Since, derivative and expectation are linear operators, we can interchange their position. Thus, we can show that

$$\begin{aligned}\frac{\partial\Sigma_f}{\partial\lambda_m} &= E\left[2\|\mathbf{u}\|^2|\tilde{u}(m)|^2 - 3|\tilde{u}(m)|^2\right]\\ &= E\Big[2\lambda_m|\tilde{u}(m)|^4 + 2\sum_{k=1,k\neq m}^{M}\lambda_k|\tilde{u}(k)|^2|\tilde{u}(m)|^2\\ &\qquad -3|\tilde{u}(m)|^2\Big]\end{aligned} \tag{15}$$

Now, using the facts $E[|\tilde{u}(m)|^2] = 1$, $E[|\tilde{u}(m)|^4] = 2E[|\tilde{u}(m)|^2]$, and $E[|\tilde{u}(m)|^2|\tilde{u}(k)|^2] = E[|\tilde{u}(m)|^2]E[|\tilde{u}(k)|^2]$ for $m \neq k$, we obtain

$$\frac{\partial\Sigma_f}{\partial\lambda_m} = \left[4\lambda_m + 2\sum_{k=1,k\neq m}^{M}\lambda_k - 3\right] \tag{16}$$

Next, using the fact $\sum_{k=1}^{M}\lambda_k = M$, we add and subtract λ_m from the middle term to get

$$\frac{\partial\Sigma_f}{\partial\lambda_i} = 4\lambda_m + 2(M-\lambda_m) - 3 \tag{17}$$

Now, using (17) and *Lemma 1*, we obtain the following inequality to test the Schur-convexity of the $\zeta_{NLMS}((\boldsymbol{\lambda}))$

$$(\lambda_1-\lambda_2)\left(\frac{\partial\Sigma_f}{\partial\lambda_1}-\frac{\partial\Sigma_f}{\partial\lambda_2}\right) = 2(\lambda_1-\lambda_2)^2\overset{?}{\geq}0. \tag{18}$$

It can be easily observed that $(\lambda_1-\lambda_2)^2 > 0$ for both $\lambda_1 > \lambda_2$ and $\lambda_1 < \lambda_2$ and hence the inequality is always satisfied. Moreover, the inequality is also satisfied if $\lambda_1 = \lambda_2$ as $(\lambda_1 - \lambda_2) = 0$ and hence confirming that the $\zeta_{NLMS}((\boldsymbol{\lambda}))$ Schur-convex.

5. SIMULATION RESULTS

In this section, we provide numerical results to support our findings. First, we demonstrate the inadequacy of eigenvalue-spread to precisely capture the the spread in input signal. Subsequently, we provide numerical examples to demonstrate the Schur-convexity of the EMSE for the LMS and NLMS.

5.1. Experiment 1: Eigenvalue-Spread

We simulate an LMS adaptive filter with tap length $M = 5$ for 1500 iterations. The signal-to-noise ratio SNR is kept fixed at 20dB. The optimal weight vector is $\boldsymbol{w}^o = [0.227,\ 0.460,\ 0.688,\ 0.460,\ 0.227]^T$. The taps of the initial estimate (i.e., $\boldsymbol{w}_{-1}$) are chosen to be all zeros. The filter is excited with two distinct complex Gaussian regressors with eigenvalues $\boldsymbol{\lambda}_1 = [50,\ 37.75,\ 25.5,\ 13.25,\ 1]^T$ and $\boldsymbol{\lambda}_2 = [50,\ 1,\ 1,\ ,1,\ 1]^T$ respectively. Note that the two regressors have the same eigenvalue-spread i.e., $\rho_1 = \rho_2 = 50$. The algorithm is run with step-size $\mu = 0.005$ and the learning curves are averaged over 500 independent runs. The results of this experiment are shown in Fig. 1.

It is clear from the results in Fig. 1 that even though the two regressors have the same eigenvalue-spread, there is a significant gap (approximately 5dB) in the MSE performance of the LMS algorithm.

5.2. Experiment 2: Majorization and Schur-Convexity

In this experiment, we analyze the effect of majorization on the MSE of LMS and NLMS algorithms. Specifically, we select four sets of eigenvalues such that $\boldsymbol{\lambda}_1 \succ \boldsymbol{\lambda}_2 \succ \boldsymbol{\lambda}_3 \succ \boldsymbol{\lambda}_4$ (see

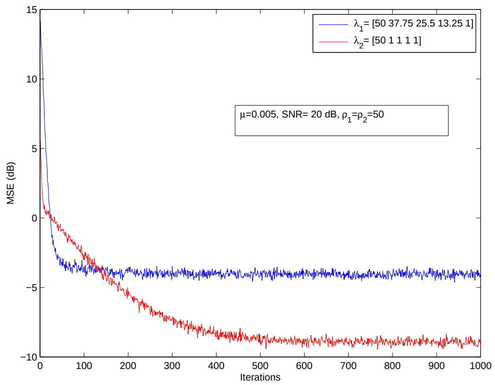

Fig. 1. MSE performance of LMS algorithm for two inputs with eigenvalue-spread $\rho_1 = \rho_2 = 50$ and set of eigenvalues $\boldsymbol{\lambda}_1$ and $\boldsymbol{\lambda}_2$.

Table. 1) and study the respective MSE. We do so by using a filter tap length $M = 5$, the step-size $\mu = 0.2$ for LMS and $\mu = 1$ for NLMS and SNR= 20dB. The steady-state values of MSE (obtained by averaging last 100 values from the MSE via simulations) are highlighted in Table. 1 and simulation results - obtained by averaging over 500 independent runs - are shown in Fig. 2.

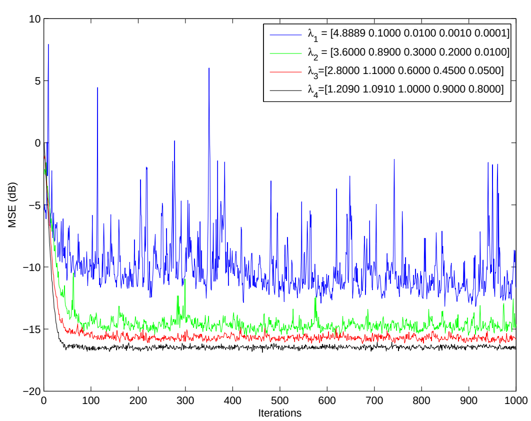

Fig. 2. MSE performance of LMS algorithm for four inputs with set of eigenvalues $\boldsymbol{\lambda}_1 \succ \boldsymbol{\lambda}_2 \succ \boldsymbol{\lambda}_3 \succ \boldsymbol{\lambda}_4$ as given in Table. 1.

From the values of the MSE given in Table. 1, we note that MSE of both the LMS and the NLMS algorithms preserve the order of majorization i.e., $\zeta(\boldsymbol{\lambda}_1) \geq \zeta(\boldsymbol{\lambda}_2) \geq \zeta(\boldsymbol{\lambda}_3) \geq \zeta(\boldsymbol{\lambda}_4)$, as expected. Further, we can see from the simulation results (in Fig. 2 and Fig. 3) that as the spread in input increases, not only the steady-state MSE increases, but variation in the ensemble averaged performance (around the stead-state MSE) also increases.

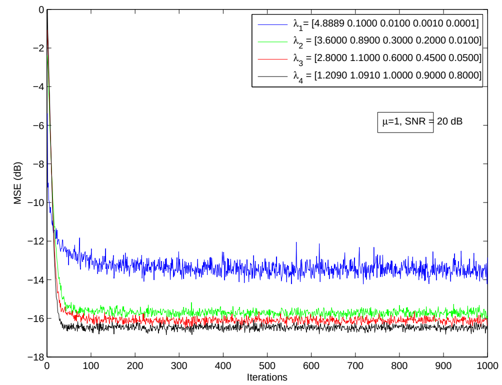

Fig. 3. MSE performance of NLMS algorithm for four inputs with set of eigenvalues $\boldsymbol{\lambda}_1 \succ \boldsymbol{\lambda}_2 \succ \boldsymbol{\lambda}_3 \succ \boldsymbol{\lambda}_4$ as given in Table. 1.

6. CONCLUSION

It is well known that input correlation is the main factor that effects the performance of steepest descent adaptive algorithms and their stochastic gradient versions. Specifically, we investigated the impact of input- spread on the MSE performance of the LMS and the NLMS algorithms and demonstrated that the condition number - that is usually used to order regressors - is not the best predictor of MSE performance. Rather, ordering regressors using the majorization concept is a more accurate measure of the spread. Moreover, it turns out the MSE of both the LMS and the NLMS are Schur-convex, which means that majorization also reflects the MSE performance order. By investigating the learning curves, one also notices that majorization not only predicts the steady state performance but also says a lot about performance of the learning curves throughout most of the adaptation process. To the best of our knowledge, this is the first work that connects adaptive filtering performance with the concepts of majorization and Schur convexity.

Eigenvalues	ζ_{LMS}	ζ_{NLMS}
$\boldsymbol{\lambda}1 = [4.4445,\ 0.5000,\ 0.0500,\ 0.0050,\ 0.0005]$	-12.2	-13.75
$\boldsymbol{\lambda}2 = [3.6000,\ 0.8900,\ 0.3000,\ 0.2000,\ 0.0100]$	-15.0	-15.85
$\boldsymbol{\lambda}3 = [2.8000,\ 1.1000,\ 0.6000,\ 0.4500,\ 0.0500]$	-15.5	-16.05
$\boldsymbol{\lambda}4 = [1.2090,\ 1.0910,\ 1.0000,\ 0.9000,\ 0.8000]$	-16.8	-16.30

Table 1. MSE Performance (in dB) of the LMS and the NLMS algorithms for the sets of eigenvalues $\boldsymbol{\lambda}$s that follow the majorization order $\boldsymbol{\lambda}_1 \succ \boldsymbol{\lambda}_2 \succ \boldsymbol{\lambda}_3 \succ \boldsymbol{\lambda}_4$.

7. REFERENCES

[1] Sayed, A. H.: ‘Adaptive filters’, *John Wiley & Sons*, 2008.

[2] Al-Naffouri, T. Y. and Sayed, A. H.: ’Transient analysis

of data-normalized adaptive filters', *IEEE Trans. Signal Process.*, vol. 51, pp.639-652, Mar. 2003.

[3] N. J. Bershad, "Analysis of the Normalized LMS Algorithm with Gaussian Inputs," *IEEE Transactions on Acoustic Speech Signal Processing*, vol ASSP-34, No. 5, pp. 793806, August, 1986.

[4] S. C. Douglas and T. H. Y. Meng, "Normalized Data Nonlinearities for LMS Adaptation," *IEEE Transactions on Signal Processing*, vol 42, No. 6, pp. 1352-1365, June, 1994.

[5] G. Barrault, M. H. Costa, J. C. M. Bermudez and A. Lenzi, "A new analytical model for the NLMS algorithm," *Proc. IEEE ICASSP*, vol. 4, pp. 41-44, March, 2005.

[6] E. M. Lobato, O. J. Tobias, and R. Seara, "Stochastic model for the NLMS algorithm with correlated Gaussian data," *Proc. IEEE ICASSP-2006*, vol. 3, pp. 760-763. 2006.

[7] M. H. Costa and J. C. M. Bermudaz, "An improved model for the normalized LMS algorithm with gaussian inputs and large number of coeeficients," *IEEE ICASSP, Orlando, FL*, vol. II, pp. 1385-1388, 2002.

[8] D. T. M. Slock, "On the convergence behavior of the LMS and the normalized LMS algorithms," *IEEE Transactions on Signal Processing*, vol 41, No. 9, pp. 2811-2825, Sept., 1993.

[9] Marshall, A. W., Olkin, I. and Arnold, B.: 'Inequalities: theory of majorization and its applications', *Springer Science & Business Media*, 2010.