Hindawi Publishing Corporation
EURASIP Journal on Advances in Signal Processing
Volume 2010, Article ID 893809, 10 pages
doi:10.1155/2010/893809

Research Article

Convergence Analysis of a Mixed Controlled $l_2 - l_p$ Adaptive Algorithm

Abdelmalek Zidouri

Electrical Engineering Department, King Fahd University of Petroleum and Minerals, Dhahran 31261, Saudi Arabia

Correspondence should be addressed to Abdelmalek Zidouri, malek@kfupm.edu.sa

Received 17 June 2010; Accepted 26 October 2010

Academic Editor: Azzedine Zerguine

A newly developed adaptive scheme for system identification is proposed. The proposed algorithm is a mixture of two norms, namely, the l_2-norm and the l_p-norm ($p \geq 1$), where a controlling parameter in the range $[0, 1]$ is used to control the mixture of the two norms. Existing algorithms based on mixed norm can be considered as a special case of the proposed algorithm. Therefore, our algorithm can be seen as a generalization to these algorithms. The derivation of the algorithm and its convexity property are reported and detailed. Also, the first moment behaviour as well as the second moment behaviour of the weights is studied. Bounds for the step size on the convergence of the proposed algorithm are derived, and the steady-state analysis is carried out. Finally, simulation results are performed and are found to corroborate with the theory developed.

1. Introduction

The least mean square (LMS) algorithm [1] is one of the most widely used adaptive schemes. Several works have been presented using the LMS or its variants [2–14], such as signed LMS [8], the least mean fourth (LMF) algorithm and its variants [15], or the mixed LMS-LMF [16–18] all of which are intuitively motivated.

The LMS algorithm is optimum only if the noise statistics are Gaussian. However, if these statistics are different from Gaussian, other criteria, such as l_p-norm ($p \neq 2$), perform better than the LMS algorithm. An alternative to the LMS algorithm which performs well when the noise statistics are not Gaussian is the LMF algorithm. A further improvement is possible when using a mixture of both algorithms, that is, the LMS and the LMF algorithms [16].

In this respect, existing algorithms based on mixed-norm (MN) criteria have been used in system identification behaving robustly in Gaussian and non-Gaussian environments. These algorithms are based on a fixed combination of the LMS and the LMF algorithms or a time varying combination of them. The time variation is used in adapting the mixed control parameter to compensate for nonstationarities and time-varying environments. The combination of error norms governed by a mixture parameter is introduced to yield a better performance than algorithms derived from a single error norm. Very attractive results are found through the use of mixed-norm algorithms [16–18]. These are based on the minimization of a mixed norm cost function in a controlled fashion, that is [16–18],

$$J_n = \alpha E[e_n^2] + (1 - \alpha) E[e_n^4], \tag{1}$$

where the error is defined as

$$e_n = d_n + w_n - \mathbf{c}_n^T \mathbf{x}_n, \tag{2}$$

d_n is the desired value, $\mathbf{c}_n$ is the filter coefficient of the adaptive filter, $\mathbf{x}_n$ is the input vector, w_n is the additive noise, and α is the mixing parameter between zero and one and set in this range to preserve the unimodal character of the cost function. It is clear from (1) that if $\alpha = 1$ the algorithm reduces to the LMS algorithm; if, however, $\alpha = 0$ the algorithm is the LMF. A careful choice for α in the interval (0,1) will enhance the performance of the algorithm. The algorithm for adjusting the tap coefficients, $\mathbf{c}_n$, is given by the following recursion:

$$\mathbf{c}_{n+1} = \mathbf{c}_n + \mu \{\alpha + 2(1 - \alpha) e_n^2\} e_n \mathbf{x}_n. \tag{3}$$

Adaptive filter algorithms designed through the minimization of equation (1) have a disadvantage when the absolute value of the error is greater than one. This makes the algorithm go unstable unless either a small value of the step size or a large value of the controlling parameter is chosen such that this unwanted instability is eliminated. Unfortunately, a small value of the step size will make the algorithm converge very slowly, and a large value of the controlling parameter will make the LMS algorithm essentially dominant.

The rest of the paper is organized as follows. In Section 2, the description of the proposed algorithm is addressed, while Section 3 deals with the convergence analysis. Section 4 details the derivation of the excess mean-square-error. The simulation results are reported in Section 5, and finally Section 6 concludes the main findings of the paper and outlines possible further work.

2. Proposed Algorithm

To overcome the above-mentioned problem, a modified approach is proposed where both constraints of the step size and the control parameter are eliminated. The proposed criterion consists of the cost function (1) where the l_p-norm is substituted for the l_4-norm. Ultimately, this should eliminate the instability in the l_4-norm and retains the good features of (1), that is, the mixed nature of the criterion if $p < 4$. The proposed scheme is defined as,

$$J_n = \alpha E[e_n^2] + (1-\alpha)E[|e_n|^p], \quad p \geq 1. \tag{4}$$

If $p = 2$, the cost function defined by (4) reduces to the LMS algorithm whatever the value of α in the range $[0, 1]$ for which the unimodality of the cost function is preserved.

For $\alpha = 0$, the algorithm reduces to the l_p-norm adaptive algorithm, and moreover if $p = 1$ results in the familiar signed LMS algorithm [14].

The value range of the lower-order p is selected to be $[1, 2]$ because

(1) for $p > 2$, the cost function may easily become large valued when the magnitude of the output error $e_n \gg 1$, leading to a potentially considerable enhancement of noise, and

(2) for $p < 1$, the gradient decreases in a positive direction, resulting in an obviously undesirable attribute for being used as a cost function. Setting the value of p within the range $[1, 2]$ provides a situation where the gradient at $e_n \gg 1$ is very much lower than that for the cases with $p = 2$. This means that the resulting algorithm can be less sensitive to noise.

For $p < 2$, l_p gives less weight for larger error and this tends to reduce the influence of aberrant noise, while it gives relatively larger weight to smaller errors and this will improve the tracking capability of the algorithm [19].

2.1. Convex Property of Cost Function. The cost function $J(c) = \alpha E[e_n^2] + (1-\alpha)E[|e_n|^p]$ is a convex function defined on $\mathbf{R}^{(N_1+N_2)}$ for $p \geq 1$, where N_1 and N_2 are the dimensions of $\mathbf{c}_1$ and $\mathbf{c}_2$, respectively.

Proof.

$$\begin{aligned}
&\alpha \left| \hat{y}_n - \mathbf{x}_n^T[a\mathbf{c}_1 + (1-a)\mathbf{c}_2] \right|^2 \\
&\quad + (1-\alpha) \left| \hat{y}_n - \mathbf{x}_n^T[a\mathbf{c}_1 + (1-a)\mathbf{c}_2] \right|^p \\
&= \alpha \left| a\left[\hat{y}_n - \mathbf{x}_n^T\mathbf{c}_1\right] + (1-a)\left[\hat{y}_n - \mathbf{x}_n^T\mathbf{c}_2\right] \right|^2 \\
&\quad + (1-\alpha) \left| a\left[\hat{y}_n - \mathbf{x}_n^T\mathbf{c}_1\right] + (1-a)\left[\hat{y}_n - \mathbf{x}_n^T\mathbf{c}_2\right] \right|^p \\
&\leq a\left\{\alpha\left[\hat{y}_n - \mathbf{x}_n^T\mathbf{c}_1\right]^2 + (1-\alpha)\left[\hat{y}_n - \mathbf{x}_n^T\mathbf{c}_1\right]^p\right\} + (1-a) \\
&\quad \times \left\{\alpha\left[\hat{y}_n - \mathbf{x}_n^T\mathbf{c}_1\right]^2 + (1-\alpha)\left[\hat{y}_n - \mathbf{x}_n^T\mathbf{c}_1\right]^p\right\}, \quad p \geq 1.
\end{aligned} \tag{5}$$

Let $f_{\hat{y}x}(\hat{y}_n, x_n)$ be the joint probability density function of $\hat{y}_n$ and $\mathbf{x}_n$. Taking the expectation value of the above, after multiplying its both sides by $f_{\hat{y}\mathbf{x}}(\hat{y}_n, \mathbf{x}_n)$, one obtains the following:

$$J(a\mathbf{c}_1 + (1-a)\mathbf{c}_2) \leq aJ(\mathbf{c}_1) + (1-a)J(\mathbf{c}_2). \tag{6}$$

This shows that the cost function J is convex. □

2.2. Analysis of the Error Surface

Case 1. Let the input autocorrelation matrix be $\mathbf{R} = E[\mathbf{x}_n\mathbf{x}_n^T]$, and the cross-correlation vector that describes the cross-correlation between the received signal (x_n) and the desired data (d_n) $\mathbf{p} = E[\mathbf{x}_n d_n]$. The error function can be more conveniently expressed as follows:

$$J_n = \sigma_x^2 - 2\mathbf{c}_n^T\mathbf{p} + \mathbf{c}_n^T\mathbf{R}\mathbf{c}_n. \tag{7}$$

It is clear from (7) that the mean-square-error (MSE) is precisely a quadratic function of the components of the tap coefficients, and the shape associated with it is hyperparaboloid. The adaptive process continuously adjusts the tap coefficients, seeking the bottom of this hyperparaboloid.

$$\mathbf{c}_{\text{opt}} = \mathbf{R}^{-1}\mathbf{p}. \tag{8}$$

Case 2. It can be shown as well that the error function for the feedback section will have a global minimum since the latter one is a convex function. As in the feedforward section, the adaptive process will continuously seek the bottom of the error function of the feedback section.

2.3. The Updating Scheme. The updating scheme is given by,

$$\mathbf{c}_{n+1} = \mathbf{c}_n + \mu\left[\alpha e_n + p(1-\alpha)|e_n|^{(p-1)}\operatorname{sign}(e_n)\right]\mathbf{x}_n, \tag{9}$$

and sufficient condition for convergence in the mean of the proposed algorithm can be shown to be given by:

$$0 < \mu < \frac{2}{\left\{\alpha + p(p-1)(1-\alpha)E\left[|w_n|^{p-2}\right]\right\}\operatorname{tr}\{\mathbf{R}\}}, \tag{10}$$

where $\text{tr}\{\mathbf{R}\}$ is the trace operation of the autocorrelation matrix $\mathbf{R}$.

In general, the step size is chosen small enough to ensure convergence of the iterative procedure and produce less misadjustment error.

3. Convergence Analysis

In this section, the convergence analysis of the proposed algorithm is detailed. The following assumptions which are quite similar to what is usually assumed in literature and which can also be justified in several practical instances are used during the conver thegence analysis of the mixed controlled $l_2 - l_p$ algorithm. For example, these are quite similar to what is usually assumed in the literature [14, 15, 20–22], and which can also be justified in several practical instances.

(A1) The input signal x_n is zero mean and having variance σ_x^2.

(A2) The noise w_n is a zero-mean independent and identically distributed process and is independent of the input signal and having zero odd moments.

(A3) The step-size is small enough for the independence assumption [14] to be valid. As a consequence, the weight-error vector is independent of the input x_n.

While assumptions (A1-A2) can be justified in several practical instances, assumption (A3) can only be attained asymptotically. The independence assumption [14] is very common in the literature and is justified in several practical instances [21]. The assumption of small step size is not necessarily true in practice but has been commonly used to simplify the analysis [14].

During the convergence analysis of the proposed algorithm only the case of $p = 1$ is considered as it is carried out for the first time. Cases for $p = 4$ can be found, for example, in [16–18].

The weight error is defined to be

$$\mathbf{v}_n = \mathbf{c}_n - \mathbf{c}_{\text{opt}}. \tag{11}$$

3.1. First Moment Behavior of the Weight Error Vector. We start by evaluating the statistical expectation of both sides of (9) which looks after subtracting c_{opt} of both sides to give

$$\mathbf{v}_{n+1} = \mathbf{v}_n + \mu[\alpha e_n + (1-\alpha)\,\text{sign}(e_n)]\mathbf{x}_n. \tag{12}$$

After substituting the error e_n defined by (2) in the above equation and taking the expectation of its both sides, this results in:

$$E[\mathbf{v}_{n+1}] = [\mathbf{I} - \alpha\mu\mathbf{R}]E[\mathbf{v}_n] + \mu(1-\alpha)E[\mathbf{x}_n\,\text{sign}(e_n)]. \tag{13}$$

Here at this point, we have to evaluate the expression $E[\text{sign}(e_n)\mathbf{x}_n]$ using Price's theorem [20] in the following way:

$$\begin{aligned} E[\mathbf{x}_n\,\text{sign}(e_n)] &= \sqrt{\frac{2}{\pi}}\frac{1}{\sigma_n}E[e_n\mathbf{x}_n] \\ &= \sqrt{\frac{2}{\pi}}\frac{1}{\sigma_n}E\left[w_n\mathbf{x}_n - \mathbf{x}_n\mathbf{x}_n^T\mathbf{v}_n\right] \\ &= -\sqrt{\frac{2}{\pi}}\frac{1}{\sigma_n}\mathbf{R}E[\mathbf{v}_n]; \end{aligned} \tag{14}$$

note that in the second step of this equation the error e_n has been substituted.

Now, we are ready to evaluate expression (13), and it is given by,

$$E[\mathbf{v}_{n+1}] = \left\{\mathbf{I} - \mu\left[\alpha + (1-\alpha)\sqrt{\frac{2}{\pi}}\frac{1}{\sigma_n}\right]\mathbf{R}\right\}E[\mathbf{v}_n]. \tag{15}$$

It is to show that the mis-alignment vector will converge to the zero vector if the step-size, μ, is given by

$$0 < \mu < \frac{2}{\left[\alpha + (1-\alpha)\sqrt{(2/\pi)}(1/\sigma_n)\right]\text{tr}\{\mathbf{R}\}}. \tag{16}$$

A more restrictive, but sufficient and simpler, condition for convergence of (12) in the mean is

$$0 < \mu < \frac{2}{\left[\alpha + (1-\alpha)\sqrt{(2/\pi J_{\min})}\right]\lambda_{\max}}, \tag{17}$$

where $\lambda_{\max}$ is the largest eigenvalue of the autocorrelation matrix $\mathbf{R}$, since in general $\text{tr}\{\mathbf{R}\} \gg \lambda_{\max}$, and $J_{\min}$ is the minimum MSE.

An inspection of (16) will immediately show that if the convergence does occur, the root mean-squared estimation error σ_n at time n is such that

$$\sigma_n > \sqrt{\frac{2}{\pi}}\left[\frac{\mu(1-\alpha)\lambda_{\max}}{2 - \mu\alpha\lambda_{\max}}\right], \tag{18}$$

where the mean-square value of the estimation error can be shown to be

$$\begin{aligned} {\sigma_n}^2 &= E[e_n^2] \\ &= E\left[\left(w_n - \mathbf{v}_n^T\mathbf{x}_n\right)\left(w_n - \mathbf{v}_n^T\mathbf{x}_n\right)^T\right] \\ &= J_{\min} + E\left[\mathbf{v}_n^T\mathbf{x}_n\mathbf{x}_n^T\mathbf{v}_n\right] \\ &= J_{\min} + \text{tr}[\mathbf{R}\mathbf{K}_n]. \end{aligned} \tag{19}$$

(a) Discussion. It can be seen from (18) that, a sufficient condition for the algorithm to converge in the mean, the following must hold:

$$0 < \mu < \frac{2}{\alpha\lambda_{\max}}. \tag{20}$$

Consequently, when $\alpha = 1$, the convergence for the LMS algorithm is proved.

3.2. Second Moment Behavior of the Weight Error Vector. From (12) we get the following expression for $\mathbf{v}_{n+1}\mathbf{v}_{n+1}^T$:

$$\mathbf{v}_{n+1}\mathbf{v}_{n+1}^T = \mathbf{v}_n\mathbf{v}_n^T + \mu[\alpha e_n + (1-\alpha)\operatorname{sign}(e_n)]\left[\mathbf{v}_n\mathbf{x}_n^T + \mathbf{x}_n\mathbf{v}_n^T\right] + \mu^2\left[\alpha^2 e_n^2 + 2\alpha(1-\alpha)|e_n| + (1-\alpha)^2\right]\mathbf{x}_n\mathbf{x}_n^T. \tag{21}$$

Let $\mathbf{K}_n = E[\mathbf{v}_n\mathbf{v}_n^T]$ define the second moment of the misalignment vector therefore, the above equation becomes, after taking the expectation of both of its sides, the following:

$$\begin{aligned}\mathbf{K}_{n+1} = \mathbf{K}_n &+ \mu\alpha\left\{E\left[\mathbf{v}_n\mathbf{x}_n^T e_n\right] + E\left[\mathbf{x}_n\mathbf{v}_n^T e_n\right]\right\} \\ &+ \mu(1-\alpha)\left\{E\left[\mathbf{v}_n\mathbf{x}_n^T \operatorname{sign}(e_n)\right] + E\left[\mathbf{x}_n\mathbf{v}_n^T \operatorname{sign}(e_n)\right]\right\} \\ &+ \mu^2\left\{\alpha^2 E\left[\mathbf{x}_n\mathbf{x}_n^T e_n^2\right] + 2\alpha(1-\alpha)E\left[\mathbf{x}_n\mathbf{x}_n^T|e_n|\right] \right. \\ &\left. +(1-\alpha)^2\mathbf{R}\right\}.\end{aligned} \tag{22}$$

Before finalizing the above expression, let us evaluate the following quantities taking into account that they are Gaussian and zero mean [20]:

$$E\left[\mathbf{x}_n\mathbf{v}_n^T \operatorname{sign}(e_n)\right] = -\sqrt{\frac{2}{\pi}}\frac{1}{\sigma_n}\mathbf{R}\mathbf{K}_n, \tag{23}$$

$$E\left[\mathbf{v}_n\mathbf{x}_n^T \operatorname{sign}(e_n)\right] = -\sqrt{\frac{2}{\pi}}\frac{1}{\sigma_n}\mathbf{K}_n\mathbf{R}, \tag{24}$$

$$E\left[\mathbf{x}_n\mathbf{v}_n^T e_n\right] = -\mathbf{R}\mathbf{K}_n, \tag{25}$$

and finally,

$$E\left[\mathbf{v}_n\mathbf{x}_n^T e_n\right] = -\mathbf{K}_n\mathbf{R}. \tag{26}$$

Substituting expressions (23)–(26) in (22) results in the following:

$$\begin{aligned}\mathbf{K}_{n+1} = \mathbf{K}_n&\left\{\mathbf{I} - \mu\left[\alpha + (1-\alpha)\sqrt{\frac{2}{\pi}}\frac{1}{\sigma_n}\right]\mathbf{R}\right\} \\ &+ \mu^2\mathbf{R}\left\{(1-\alpha)^2 + \left[\alpha^2 - 2\alpha(1-\alpha)\sqrt{\frac{2}{\pi}}\frac{1}{\sigma_n}\right]\right. \\ &\left.\times[J_{\min} + \operatorname{tr}(\mathbf{R}\mathbf{K}_n)]\right\} \\ &- \mu\left[\alpha + (1-\alpha)\sqrt{\frac{2}{\pi}}\frac{1}{\sigma_n}\right]\mathbf{R}K_n.\end{aligned} \tag{27}$$

During the derivation of the above equation, expressions $E[\mathbf{x}_n\mathbf{x}_n^T e_n^2]$ and $E[\mathbf{x}_n\mathbf{x}_n^T|e_n|]$ are evaluated, respectively, as follows:

$$\begin{aligned}E\left[\mathbf{x}_n\mathbf{x}_n^T e_n^2\right] &= E\left[\mathbf{x}_n\mathbf{x}_n^T\left(\omega_n - \mathbf{v}_n^T\mathbf{x}_n\right)^2\right] \\ &= \mathbf{R}\{J_{\min} + \operatorname{tr}[\mathbf{R}\mathbf{K}_n]\},\end{aligned} \tag{28}$$

and

$$\begin{aligned}E\left[\mathbf{x}_n\mathbf{x}_n^T|e_n|\right] &= E\left[\mathbf{x}_n\mathbf{x}_n^T e_n \operatorname{sign}(e_n)\right] \\ &= -\sqrt{\frac{2}{\pi}}\frac{1}{\sigma_n}E\left[\mathbf{x}_n\mathbf{x}_n^T e_n^2\right] \\ &= -\sqrt{\frac{2}{\pi}}\frac{1}{\sigma_n}\mathbf{R}\{J_{\min} + \operatorname{tr}[\mathbf{R}\mathbf{K}_n]\}.\end{aligned} \tag{29}$$

Both of these expressions are substituted in (22) to result in its simplified form (27).

Now, denote by σ_∞ and $\mathbf{K}_\infty$ the limiting values of σ_n and $\mathbf{K}_n$, respectively; then closed-form expressions for the limiting (steady-state) values of the second moment matrix and error power are derived next.

It is assumed that the autocorrelation matrix, $\mathbf{R}$, is positive definite [23] with eigenvalues, λ_i; hence, it can be factorized as;

$$\mathbf{R} = \mathbf{Q}\mathbf{\Lambda}\mathbf{Q}^T, \tag{30}$$

where $\mathbf{\Lambda}$ is the diagonal matrix of eigenvalues

$$\mathbf{\Lambda} = \operatorname{diag}(\lambda_1, \lambda_2, \ldots, \lambda_N), \tag{31}$$

and Q is the orthonormal matrix whose ith column is the eigenvector of $\mathbf{R}$ associated with the ith eigenvalue, that is,

$$\mathbf{Q}^T\mathbf{Q} = \mathbf{I}, \tag{32}$$

which results in

$$\mathbf{G}_n = \mathbf{Q}^T\mathbf{K}_n\mathbf{Q}, \tag{33}$$

hence (27) can be written as

$$\begin{aligned}\mathbf{G}_{n+1} = \mathbf{G}_n&\left\{\mathbf{I} - \mu\left[\alpha + (1-\alpha)\sqrt{\frac{2}{\pi}}\frac{1}{\sigma_n}\right]\mathbf{\Lambda}\right\} \\ &+ \mu^2\mathbf{\Lambda}\left\{(1-\alpha)^2 + \left[\alpha^2 - 2\alpha(1-\alpha)\sqrt{\frac{2}{\pi}}\frac{1}{\sigma_n}\right]\right. \\ &\left.\times[J_{\min} + \operatorname{tr}(\mathbf{\Lambda}\mathbf{G}_n)]\right\} \\ &- \mu\left[\alpha + (1-\alpha)\sqrt{\frac{2}{\pi}}\frac{1}{\sigma_n}\right]\mathbf{\Lambda}\mathbf{G}_n.\end{aligned} \tag{34}$$

We are now ready to decompose the above matrix equation into its scalar form as:

$$\begin{aligned}g_{n+1}^{i,j} = &\left\{1 - \mu\left[\alpha + (1-\alpha)\sqrt{\frac{2}{\pi}}\frac{1}{\sigma_n}\right]\left(\lambda_i + \lambda_j\right)\right\}g_n^{i,j} \\ &+ \mu^2\lambda_i\left\{(1-\alpha)^2 + \left[\alpha^2 - 2\alpha(1-\alpha)\sqrt{\frac{2}{\pi}}\frac{1}{\sigma_n}\right]\right. \\ &\left.\times\left[J_{\min} + \sum_{i=1}^{N}\lambda_i g_n^{i,i}\right]\right\}\delta_{i,j},\end{aligned} \tag{35}$$

where

$$\delta_{i,j} = \begin{cases} 1 & \text{if } i = j, \\ 0, & \text{otherwise,} \end{cases} \quad (36)$$

and $g_n^{i,j}$ is the (i, j)th scalar element of the matrix $\mathbf{G}_n$.

Two cases can be considered for the step size μ so that the weight vector converges in the mean square sense.

(1) Case $i \neq j$. In this case, (35) consists of the off-diagonal elements of matrix G_n and will look like the following:

$$g_{n+1}^{i,j} = \left\{1 - \mu\left[\alpha + (1-\alpha)\sqrt{\frac{2}{\pi}\frac{1}{\sigma_n}}\right]\left(\lambda_i + \lambda_j\right)\right\}g_n^{i,j}; \quad (37)$$

consequently, the range of the step size parameter is dictated by

$$0 < \mu < \frac{2}{\left[\alpha + (1-\alpha)\sqrt{(2/\pi)}(1/\sigma_n)\right]\left[\lambda_i + \lambda_j\right]}. \quad (38)$$

As it was in the case of the mean convergence, a sufficient condition for mean square convergence is

$$0 < \mu < \frac{1}{\left[\alpha + (1-\alpha)\sqrt{2/\pi J_{\min}}\right] \operatorname{tr}\{\mathbf{R}\}}. \quad (39)$$

(2) Case $i = j$. In this case, (35) consists of only the diagonal elements of matrix G_n and will look like the following:

$$\begin{aligned} g_{n+1}^{i,i} = &\left\{1 - 2\mu\left[\alpha + (1-\alpha)\sqrt{\frac{2}{\pi}\frac{1}{\sigma_n}}\right]\lambda_i \right.\\ &\left. + \mu^2\left[\alpha^2 - 2\alpha(1-\alpha)\sqrt{\frac{2}{\pi}\frac{1}{\sigma_n}}\right]\lambda_i^2\right\}g_n^{i,i} \\ &+ \mu^2\lambda_i\left\{(1-\alpha)^2 + \left[\alpha^2 - 2\alpha(1-\alpha)\sqrt{\frac{2}{\pi}\frac{1}{\sigma_n}}\right]\right. \\ &\left.\times\left[J_{\min} + \sum_{j=1, j\neq i}^{N}\lambda_j g_n^{j,j}\right]\right\}; \end{aligned} \quad (40)$$

correspondingly, the range of the step size parameter for convergence in the mean square sense is given by

$$0 < \mu < \frac{2\left[\alpha + (1-\alpha)\sqrt{2/\pi}(1/\sigma_n)\right]}{\left[\alpha^2 - 2\alpha(1-\alpha)\sqrt{\sqrt{2/\pi}}(1/\sigma_n)\right]\lambda_i}. \quad (41)$$

(b) Discussion. Note that $\alpha = 0$ will result in zero in the denominator of expression (41) and therefore will make μ take any value in the range of positive numbers, a contradiction with the ranges of values for the step sizes of LMS and LMF algorithms. Moreover, any value for α in $]0, 1]$ will make of the step size μ set by (41) less than zero, also this condition is discarded. This concludes that it is safer to use the more realistic bounds of (39) which will guarantee stability regardless of the value of α, and therefore will be considered here.

Once again, it is easy to see that if the convergence in the mean-square occurs, consequently the following occurs

$$\sigma_n > \sqrt{\frac{2}{\pi}\left[\frac{\mu(1-\alpha)\lambda_{\max}}{1-\mu\alpha\lambda_{\max}}\right]}. \quad (42)$$

4. Derivation of the Excess Mean-Square-Error (EMSE)

In this section, the derivation of the EMSE will be performed for the general case of p. First, let us define the *a priori* estimation error e_{an}

$$e_{an} = \mathbf{v}_{n+1}^T\mathbf{x}_n. \quad (43)$$

Second, the following assumption is to be used in the following ensuing analysis:

(A4) The *a priori* estimation error e_{an} with zero-mean is independent of $\{w_n\}$.

The updating scheme of the proposed algorithm defined in (9) can be set up into the following recursion:

$$\mathbf{c}_{n+1} = \mathbf{c}_n + \mu g(e_n)\mathbf{x}_n, \quad (44)$$

where the error function $g(e_n)$ is given by

$$g(e_n) = \alpha e_n + p\overline{\alpha}|e_n|^{p-1}\operatorname{sign}(e_n), \quad (45)$$

where $\overline{\alpha} = (1-\alpha)$.

In order to find the expression of the EMSE of the algorithm (defined as $\zeta_{\text{EMSE}} = E[e_{an}^2]$), we need to evaluate the following relation:

$$2E[e_{an}g(e_n)] = \mu\operatorname{Tr}(\mathbf{R})E\left[|g(e_n)|^2\right]. \quad (46)$$

Taking the left-hand side of (46), we can write

$$2E[e_{an}g(e_n)] = 2E\left[e_{an}\left(\alpha e_n + p\overline{\alpha}|e_n|^{p-1}\operatorname{sign}(e_n)\right)\right] \quad (47)$$

At this point, we make use of the Taylor series expansion to expand $g(e_n)$ with respect to e_n around w_n as

$$g(e_n) = g(w_n) + g_e^{(1)}(w_n)e_{an} + \frac{1}{2}g_e^{(2)}(w_n)e_{an}^2 + O(e_{an}), \quad (48)$$

where $g_e^{(1)}(w_n)$ and $g_e^{(2)}(w_n)$ are, respectively, the first-order and second-order derivatives of $g(e_n)$ with respect to e_n evaluated around w_n, and $O(e_{an})$ denotes the third, and higher-order terms of e_{an}.

Using (45), we can write

$$\begin{aligned} g_e^{(1)}(w_n) &= \alpha + p(p-1)\overline{\alpha}|w_n|^{p-2}(\operatorname{sign}(w_n))^2 \\ &\quad + p\overline{\alpha}|w_n|^{p-1}\cdot 2\delta(w_n) \\ &= \alpha + p(p-1)\overline{\alpha}|w_n|^{p-2}. \end{aligned} \quad (49)$$

Similarly, we can obtain

$$g_e^{(2)}(w_n) = p(p-1)(p-2)\overline{\alpha}|w_n|^{p-3}\operatorname{sign}(w_n) \tag{50}$$

Substituting (48) in (47) we get

$$2E[e_{an}g(e_n)] = 2E\Big[\Big(g(w_n)e_{an} + g_e^{(1)}(w_n)e_{an}^2 + O(e_{an})\Big)\Big] \tag{51}$$

Using (A4) and ignoring $O(e_{an})$, we obtain

$$2E[e_{an}g(e_n)] \approx 2E\Big[g_e^{(1)}(w_n)e_{an}^2\Big] \tag{52}$$

Using (49), we get

$$2E[e_{an}g(e_n)] = 2\Big\{\alpha + p(p-1)\overline{\alpha}E\Big[|w_n|^{p-2}\Big]\Big\}\zeta_{\text{EMSE}} \tag{53}$$

Using the Price's theorem to evaluate the expectation $E[|w_n|^{p-2}\operatorname{sign}(w_n)]$ as

$$E\Big[|w_n|^{p-2}\operatorname{sign}(w_n)\Big] = \sqrt{\frac{2}{\pi}}\frac{1}{\sigma_w}\psi_w^{p-1}, \tag{54}$$

where $E[|w_n|^p] = \psi_w^p$. So (53) becomes

$$2E[e_{an}g(e_n)] = 2\left\{\alpha + p(p-1)\overline{\alpha}\sqrt{\frac{2}{\pi}}\frac{1}{\sigma_w}\psi_w^{p-1}\right\}\zeta_{\text{EMSE}}. \tag{55}$$

Now taking the right-hand side of (46), we require $|g(e_n)|^2$. So, we write

$$|g(e_n)|^2 = \alpha^2 e_n^2 + p^2\overline{\alpha}^2|e_n|^{2p-2} + 2p\alpha\overline{\alpha}|e_n|^p\operatorname{sign}(e_n). \tag{56}$$

Therefore,

$$\begin{aligned}\mu\operatorname{Tr}(\mathbf{R})E\Big[|g(e_n)|^2\Big] &= \mu\operatorname{Tr}(\mathbf{R})E\Big[|g(w_n)|^2 + \Big|g_e^{(1)}(w_n)\Big|^2 e_{an} \\ &\quad + \frac{1}{2}\Big|g_e^{(2)}(w_n)\Big|^2 e_{an}^2 + O(e_{an})\Big]\end{aligned} \tag{57}$$

Using (A2) and (A4) and ignoring $O(e_{an})$, we write (57) as

$$\begin{aligned}\mu\operatorname{Tr}(\mathbf{R})E\Big[|g(e_n)|^2\Big] &= \mu\operatorname{Tr}(\mathbf{R})\Big[E|g(w_n)|^2 \\ &\quad + \frac{1}{2}E\Big|g_e^{(2)}(w_n)\Big|^2 e_{an}^2\Big].\end{aligned} \tag{58}$$

By using (56), we can evaluate $|g_e^{(2)}(w_n)|^2$ as

$$\begin{aligned}\Big|g_e^{(2)}(e_n)\Big|^2 &= 2\alpha^2 + (2p-2)(2p-3)p^2\overline{\alpha}^2|e_n|^{2p-4} \\ &\quad + 2p^2(p-1)\alpha\overline{\alpha}|e_n|^{p-2}\operatorname{sign}(e_n).\end{aligned} \tag{59}$$

Therefore, using (56) and (59), we can evaluate

$$\begin{aligned}&\Big[E|g(w_n)|^2 + \frac{1}{2}E\Big|g_e^{(2)}(w_n)\Big|^2 e_{an}^2\Big] \\ &= \alpha^2\sigma_w^2 + p^2\overline{\alpha}^2\psi_w^{2p-2} + 2\sqrt{\frac{2}{\pi}}\frac{1}{\sigma_w}p\alpha\overline{\alpha}\psi_w^{p+1} \\ &\quad + \Big[\alpha^2 + (p-1)(2p-3)p^2\overline{\alpha}^2\psi_w^{2p-4} \\ &\quad + \sqrt{\frac{2}{\pi}}\frac{1}{\sigma_w}p^2(p-1)\alpha\overline{\alpha}\psi_w^{p-1}\Big]\zeta_{\text{EMSE}}.\end{aligned} \tag{60}$$

Now letting

$$A = \alpha^2\sigma_w^2 + p^2\overline{\alpha}^2\psi_w^{2p-2} + 2\sqrt{\frac{2}{\pi}}\frac{1}{\sigma_w}p\alpha\overline{\alpha}\psi_w^{p+1}, \tag{61}$$

$$B = \alpha + p(p-1)\overline{\alpha}\sqrt{\frac{2}{\pi}}\frac{1}{\sigma_w}\psi_w^{p-1}, \tag{62}$$

$$\begin{aligned}C &= \alpha^2 + (p-1)(2p-3)p^2\overline{\alpha}^2\psi_w^{2p-4}, \\ &\quad + \sqrt{\frac{2}{\pi}}\frac{1}{\sigma_w}p^2(p-1)\alpha\overline{\alpha}\psi_w^{p-1},\end{aligned} \tag{63}$$

we can write (58) as

$$\mu\operatorname{Tr}(\mathbf{R})E\Big[|g(e_n)|^2\Big] = \mu\operatorname{Tr}(\mathbf{R})[A + C\zeta_{\text{EMSE}}], \tag{64}$$

and subsequently (46) can be concisely expressed as

$$2B\zeta_{\text{EMSE}} = \mu\operatorname{Tr}(\mathbf{R})[A + C\zeta_{\text{EMSE}}], \tag{65}$$

and the EMSE can be evaluated as

$$\zeta_{\text{EMSE}} = \frac{\mu A\operatorname{Tr}(\mathbf{R})}{2B - \mu C\operatorname{Tr}(\mathbf{R})}. \tag{66}$$

5. Simulation Results

In this section, the performance analysis of the proposed mixed controlled $l_2 - l_p$ adaptive algorithm is investigated in an unknown system identification problem for different values of p and different values of the mixing parameter α. The simulations reported here are based on an FIR channel system identification defined by the following channel:

$$\mathbf{c}_{\text{opt}} = [0.227, 0.460, 0.688, 0.460, 0.227]^T. \tag{67}$$

Three different noise environments have been considered namely, Gaussian, uniform, and Laplacian. The length of the adaptive filter is the same as that of the unknown system. The learning curves are obtained by averaging 600 independent runs. Two scenarios are considered for the case of the value of p, that is, $p = 1$ and $p = 4$. The performance measure considered here is the excess mean-square-error (EMSE).

Figures 2, 3, and 4 depict the convergence behavior of the proposed algorithm for different values of α in

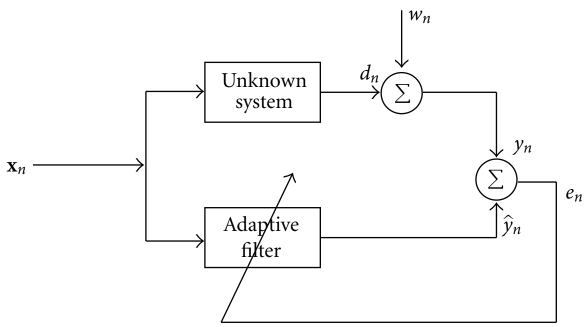

Figure 1: Block diagram representation for the proposed algorithm.

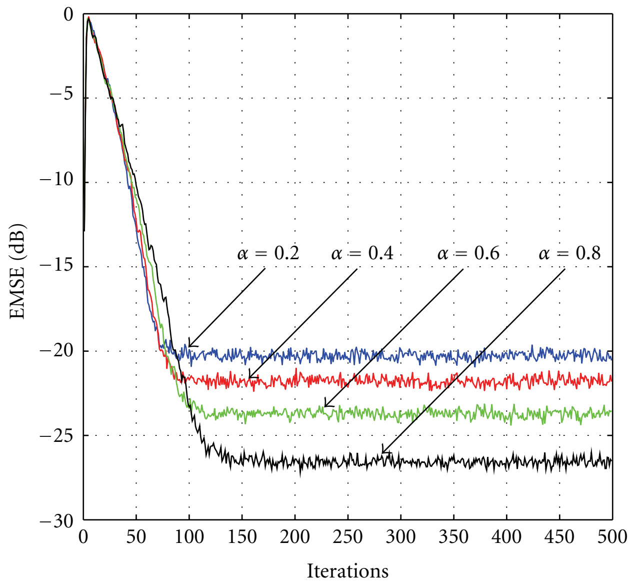

Figure 2: Effect of α on the learning curves of the proposed algorithm in an AWGN noise environment scenario for $p = 1$.

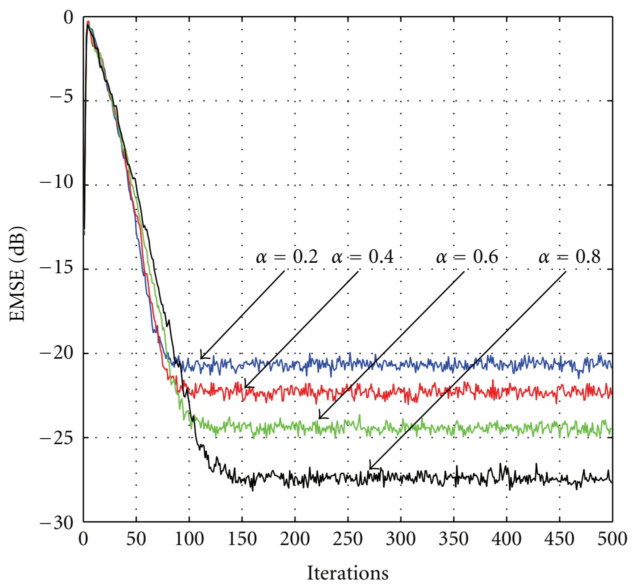

Figure 3: Effect of α on the learning curves of the proposed algorithm in a Laplacian noise environment scenario for $p = 1$.

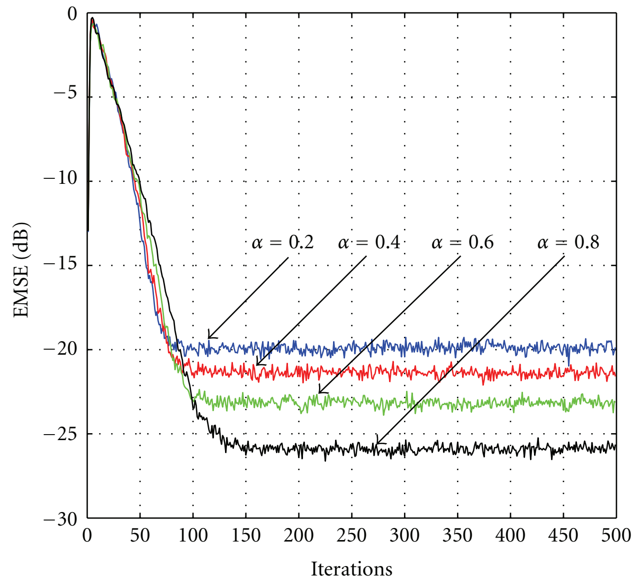

Figure 4: Effect of α on the learning curves of the proposed algorithm in a uniform noise environment scenario for $p = 1$.

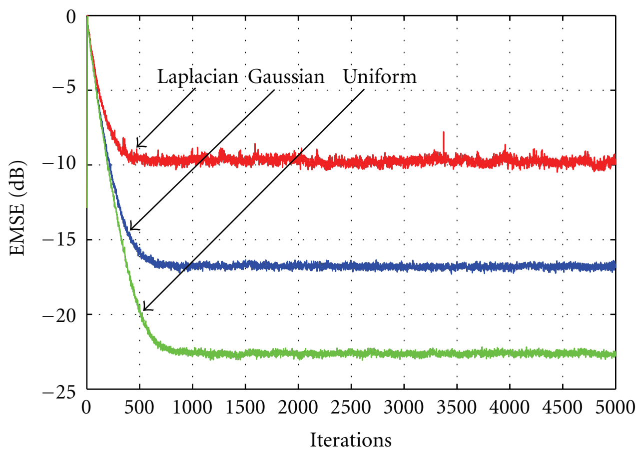

Figure 5: Learning curves of the proposed algorithm in different noise environments scenarios for $\alpha = 0.2$ and SNR of 0 dB.

a white Gaussian noise, Laplacian noise, and uniform noise, respectively, for the case of $p = 1$. As can be depicted from these figures the best performance is obtained when $\alpha = 0.8$. More importantly, the best noise statistics for this scenario is when the noise is Laplacian distributed. An enhancement in performance is obtained, and about a 2 dB improvement is achieved for all values of α. Also, one can notice that the worst performance is obtained when the noise is uniformly distributed.

Figures 5, 6, 7, 8, 9 and 10 report the performance of the proposed algorithm for an SNR of 0 dB, 10 dB and 20 dB, respectively, for the case of $p = 4$. Figures 5 and 6 are the result of the simulations for $\alpha = 0.2$ and $\alpha = 0.8$, respectively. A consistency in performance of the proposed algorithm in these scenarios for the uniform noise as far as the lowest EMSE is reached by the proposed algorithm.

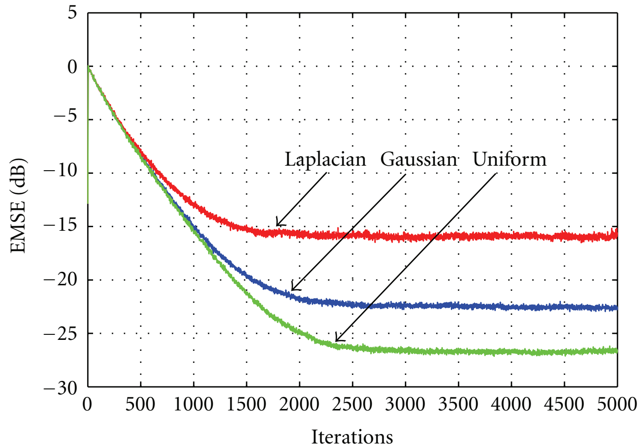

Figure 6: Learning curves of the proposed algorithm in different noise environments scenarios for $\alpha = 0.8$ and SNR of 0 dB.

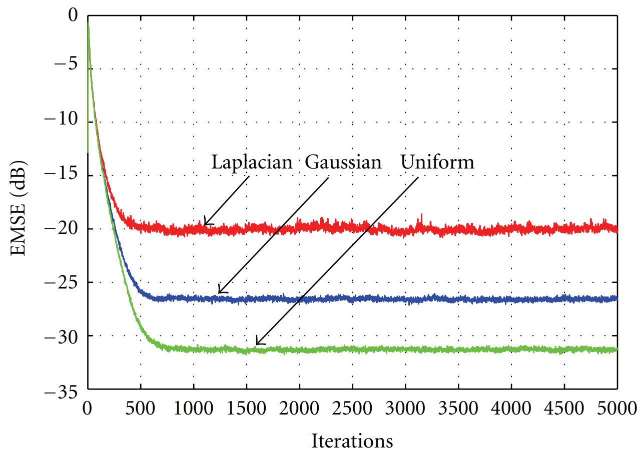

Figure 7: Learning curves of the proposed algorithm in different noise environments scenarios for $\alpha = 0.2$ and SNR of 10 dB.

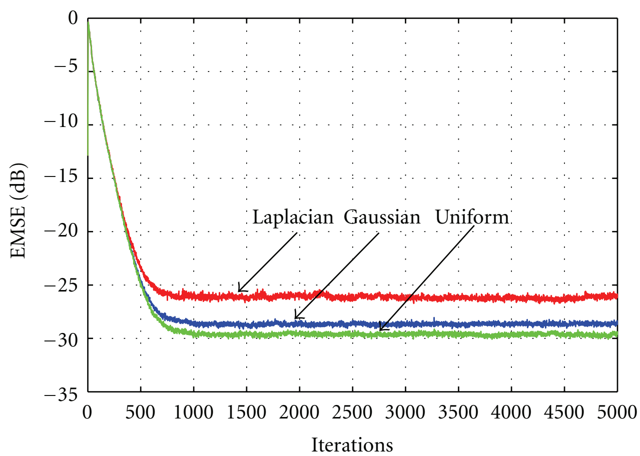

Figure 8: Learning curves of the proposed algorithm in different noise environments scenarios for $\alpha = 0.8$ and SNR of 10 dB.

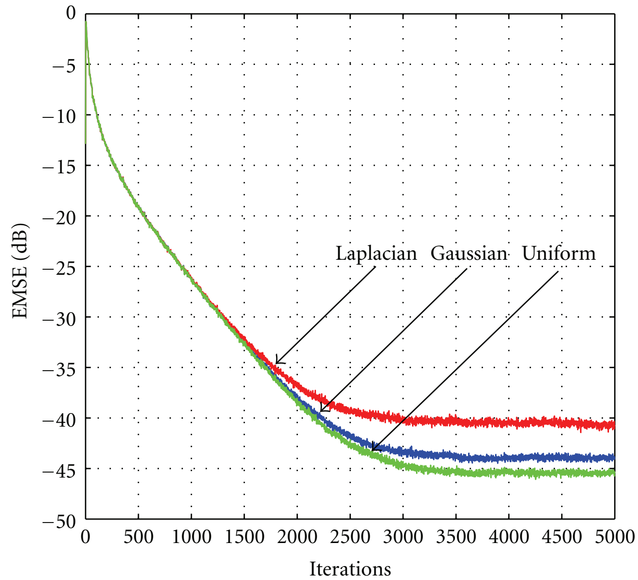

Figure 9: Learning curves of the proposed algorithm in different noise environments scenarios for $\alpha = 0.2$ and SNR of 20 dB.

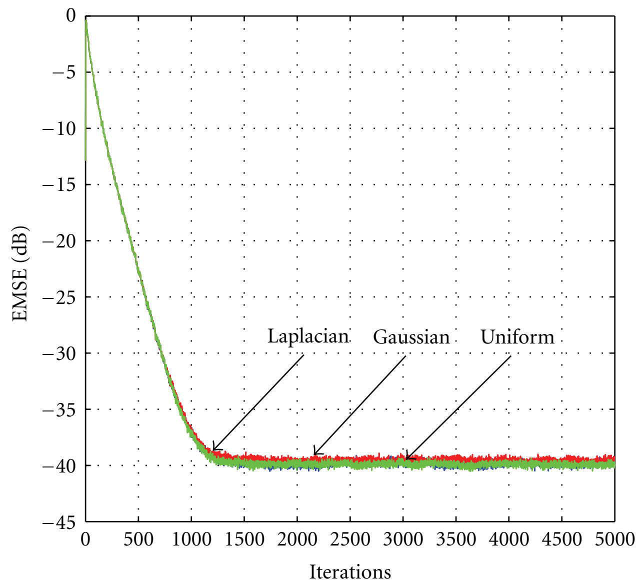

Figure 10: Learning curves of the proposed algorithm in different noise environments scenarios for $\alpha = 0.8$ and SNR of 20 dB.

Similar behaviour is obtained by the proposed algorithm in Figures 7 and 8 where Figures 7 and 8 report the simulations results of the proposed algorithm for $\alpha = 0.2$ and $\alpha = 0.8$, respectively, for an SNR of 10 dB.

In the case of an SNR of 20 dB, Figures 9 and 10 depict the results. The case of $\alpha = 0.2$ is shown in Figure 9 while that of $\alpha = 0.8$ is shown in Figure 10. One can see that, even though the proposed algorithm is still performing better in the uniform noise environment, as shown in Figure 9, for $\alpha = 0.2$, however, identical performance is obtained by the different noise environments when $\alpha = 0.8$ as reported in

TABLE 1: Theoretical and simulation EMSE for $p = 4, \alpha = 0.2$.

	Gaussian		Laplacian		Uniform	
	Theoretical	Simulation	Theoretical	Simulation	Theoretical	Simulation
0 dB	−16.9	−16.85	−9.62	−9.82	−22.81	−22.6
10 dB	−26.02	−26.53	−19.33	−19.99	−31.64	−31.29
20 dB	−44.14	−43.93	−40.34	−40.55	−45.14	−45.43

TABLE 2: Theoretical and simulation EMSE for $p = 4, \alpha = 0.8$.

	Gaussian		Laplacian		Uniform	
	Theoretical	Simulation	Theoretical	Simulation	Theoretical	Simulation
0 dB	−22.47	−22.6	−14.26	−16.02	−27.65	−26.59
10 dB	−28.7	−28.64	−26.41	−26.32	−29.15	−29.57
20 dB	−39.28	−39.87	−39.24	−39.58	−39.28	−39.92

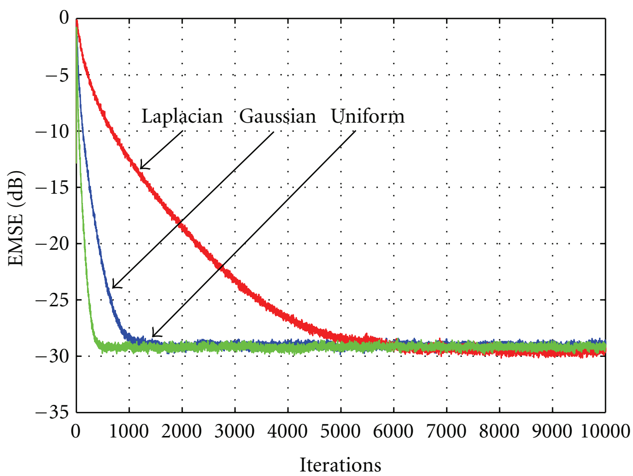

FIGURE 11: Learning behavior of the proposed algorithm in the different noise environments scenarios for $p = 4$ and $\alpha = 0.2$.

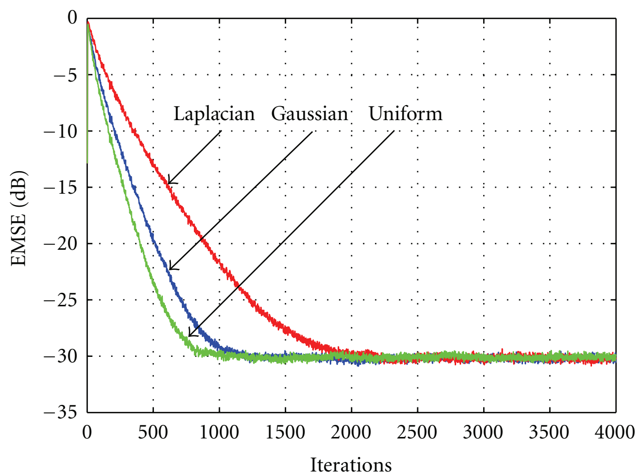

FIGURE 12: Learning behavior of the proposed algorithm in the different noise environments scenarios for $p = 4$ and $\alpha = 0.8$.

Figure 10. The theoretical findings confirm these results as will be seen later.

From the above results, one can conclude that when $\alpha = 0.2$ the proposed algorithm is biased towards the LMF algorithm, in contrast to the case when $\alpha = 0.8$, the proposed algorithm is biased towards the LMS algorithm.

Next, to assess further the performance of the proposed algorithm for the same steady-state value, two different cases for α are considered, that is, $\alpha = 0.2$ and $\alpha = 0.8$. Figures 11 and 12 illustrate the learning behavior of the proposed algorithm for $\alpha = 0.2$ and $\alpha = 0.8$, respectively, both are for $p = 4$. As can be seen from these figures that the best performance is obtained with uniform noise while the worst performance is obtained with Laplacian. The mixing variable α had little effect on the speed of convergence of the proposed algorithm when the noise is uniformly and Gaussian distributed. However, as can be seen from Figure 12 in the case of Laplacian noise, $\alpha = 0.8$ has decreased the speed of convergence of the proposed algorithm from 55000 iterations (in the case of $\alpha = 0.2$) to almost 2000 iterations. A gain of 3500 iterations in favor of the proposed algorithm when the noise is Laplacian distributed.

Finally, the analytical results for the steady-state EMSE derived for the proposed algorithm given in (66) are compared with the ones obtained from simulation for Gaussian, Laplacian, and uniform noise environments with an SNR of 0 dB, 10 dB, and 20 dB. This comparison is reported in Tables 1-2, and as can be seen from these tables, a close agreement exists between theory and the simulation results as mentioned earlier, for the case of $p = 4$ and $\alpha = 0.8$, that similar performance by the different noise environments is obtained for and SNR of 20 dB as shown in Table 2.

6. Conclusion

A new adaptive scheme for system identification has been introduced, where a controlling parameter in the range $[0, 1]$ is used to control the mixture of the two norms. The derivation of the algorithm is worked out, and the convexity property is proved for this algorithm. Existing algorithms, for example [16–18] can be considered as a special case of the proposed algorithm. Also, the first moment behaviour as well as the second moment behaviour of the weights are studied. Bounds for the step size on the convergence of the proposed algorithm are derived. Finally, the steady-state analysis was carried out; simulation results performed for the purpose of validating theory are found to be in good agreement with the theory developed.

The proposed algorithm has been applied so far to a system identification scenario, for example, echo cancellation. As a future extension, recent work is going on the application of the proposed algorithm to mitigate the effects of intersymbol interference in a communication system.

Acknowledgment

The author would like to acknowledge the support of King Fahd University of Petroleum and Minerals to carry out this research.

References

[1] B. Widrow and S. D. Stearns, *Adaptive Signal Processing*, Prentice-Hall, Englewood Cliffs, NJ, USA, 1985.

[2] S. Sherman, "Non-mean-square error criteria," *IRE Transactions on Information Theory*, vol. 4, no. 3, pp. 125–126, 1958.

[3] J. I. Nagumo and A. Noda, "A learning method for system identification," *IEEE Transactions on Automatic Control*, vol. 12, pp. 282–287, 1967.

[4] T. A. C. M. Claasen and W. F. G. Mecklenbraeuker, "Comparisons of the convergence of two algorithms for adaptive FIR digital filters," *IEEE Transactions on Circuits and Systems*, vol. 28, no. 6, pp. 510–518, 1981.

[5] A. Gersho, "Adaptive filtering with binary reinforcement," *IEEE Transactions on Information Theory*, vol. 30, no. 2, pp. 191–199, 1984.

[6] A. Feuer and E. Weinstein, "Convergence analysis of LMS filters with uncorrelated data," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 33, no. 1, pp. 222–230, 1985.

[7] N. J. Bershad, "Behavior of the e-normalized LMS algorithm with Gaussian inputs," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 35, no. 5, pp. 636–644, 1987.

[8] E. Eweda, "Convergence of the sign algorithm for adaptive filtering with correlated data," *IEEE Transactions on Information Theory*, vol. 37, no. 5, pp. 1450–1457, 1991.

[9] S. C. Douglas and T. H. Y. Meng, "Stochastic gradient adaptation under general error criteria," *IEEE Transactions on Signal Processing*, vol. 42, no. 6, pp. 1335–1351, 1994.

[10] T. Y. Al-Naffouri, A. Zerguine, and M. Bettayeb, "A unifying view of error nonlinearities in LMS adaptation," in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP '98)*, pp. 1697–1700, May 1998.

[11] H. Zhang and Y. Peng, "l_p-norm based minimisation algorithm for signal parameter estimation," *Electronics Letters*, vol. 35, no. 20, pp. 1704–1705, 1999.

[12] S. Siu and C. F. N. Cowan, "Performance analysis of the lp norm back propagation algorithm for adaptive equalisation," *IEE Proceedings, Part F: Radar and Signal Processing*, vol. 140, no. 1, pp. 43–47, 1993.

[13] R. A. Vargas and C. S. Burrus, "The direct design of recursive or IIR digital filters," in *Proceedings of the 3rd International Symposium on Communications, Control, and Signal Processing (ISCCSP '08)*, pp. 188–192, March 2008.

[14] S. Haykin, *Adaptive Filter Theory*, Prentice-Hall, Upper-Saddle River, NJ, USA, 4th edition, 2002.

[15] E. Walach and B. Widrow, "The least mean fourth (LMF) adaptive algorithm and its family," *IEEE Transactions on Information Theory*, vol. 30, no. 2, pp. 275–283, 1984.

[16] O. Tanrikulu and J. A. Chambers, "Convergence and steady-state properties of the least-mean mixed-norm (LMMN) adaptive algorithm," *IEE Proceedings Vision, Image & Signal Processing*, vol. 143, no. 3, pp. 137–142, 1996.

[17] A. Zerguine, C. F. N. Cowan, and M. Bettayeb, "LMS-LMF adaptive scheme for echo cancellation," *Electronics Letters*, vol. 32, no. 19, pp. 1776–1778, 1996.

[18] A. Zerguine, C. F. N. Cowan, and M. Bettayeb, "Adaptive echo cancellation using least mean mixed-norm algorithm," *IEEE Transactions on Signal Processing*, vol. 45, no. 5, pp. 1340–1343, 1997.

[19] S. Siu, G. J. Gibson, and C. F. N. Cowan, "Decision feedback equalisation using neural network structures and performance comparison with standard architecture," *IEE Proceedings, Part I: Communications, Speech and Vision*, vol. 137, no. 4, pp. 221–225, 1990.

[20] R. Price, "A useful theorem for non-linear devices having Gaussian inputs," *IEEE Transactions on Information Theory*, vol. 4, pp. 69–72, 1958.

[21] J. E. Mazo, "On the independence theory of equalizer convergence," *The Bell System Technical Journal*, vol. 58, no. 5, pp. 963–993, 1979.

[22] O. Macchi, *Adaptive Processing: The Least Mean Squares Approach with Applications in Transmission*, John Wiley & Sons, West Sussex, UK, 1995.

[23] A. H. Sayed, *Fundamentals of Adaptive Filtering*, Wiley-Interscience, New York, NY, USA, 2003.