HOW DOES THE CLIPPED LMS OUTPERFORM THE LMS?

M. Kallel Smaoui, Y. Ben Jemaa, and M. Jaidane

Signals and Systems Research Unit National Engineering School of Tunis
BP 37, Le Belvédère, 1002, Tunisia.
U2S@enit.rnu.tn

ABSTRACT

The clipped input LMS (known as CLMS) is a common adaptive algorithm that has a lower complexity than the conventional LMS. Furthermore, the CLMS is appreciated in real-time audio embedded systems as it uses an implicit input normalization, which is necessary for non stationary audio/speech inputs. In this paper -through an exact convergence analysis- done without the common classical assumption, we show that, the CLMS algorithm can outperform the LMS in some situations at the optimal conditions. We stress the invalid common result that LMS is faster than CLMS. This is done for finite alphabet inputs where the lower the input cardinality is, the lower is CLMS/LMS complexity.

1. INTRODUCTION

Adaptive filters are useful in a wide variety of areas such as digital communications, real-time audio embedded systems and biomedical electronics... [1]-[9]. The adaptive procedure allows real-time optimization, without a human operator and everlasting in case of variations [1].
Although the LMS is known for its simplicity and robustness, its computational cost in terms of the processor's speed is problematic when the signals have large bandwidths [7]. Therefore, past and recent works are interested in the sign adaptive filters, which allow to reduce complexity by clipping the estimation error (Sign Algorithm (SA)), the input data (Clipped Algorithm (CLMS) or Signed Regressor Algorithm (SRA) or both (Sign Sign Algorithm (SSA))[2, 3, 6, 8].

The CLMS is based on input's normalization, and is known to be less complex than the LMS. It is essentially used to reduce the computational time of certain applications [8]. The performances, particulary the convergence speed of the CLMS, were believed to be slower than those of the LMS.
For the transient behavior, Bershad [10] states that the CLMS has a convergence rate that is slightly lower than the LMS's rate. It was shown to be slower only by a factor of $\frac{2}{\pi}$ for the same steady-state mean square error.
We must note that performances of the CLMS were studied under some restrictive hypotheses. In fact, Eweda [4] has only proposed conditions of stability and convergence rates of the CLMS for correlated Gaussian input and small step size. Koike [8], however, gave a theoretical formulation for describing the convergence process of the adaptive filter using the CLMS. Results were derived for near realistic assumptions, mainly the independence one. These classical approaches do not account for finite alphabet input signals, which is unrealistic in the digital transmission context, where transmitted signals belong to a finite alphabet set.
The present work addresses an exact performance study of the CLMS algorithm. As proposed in [9], the finite alphabet approach leads to exact analysis results for various structures, algorithms and context (Identification, Equalization, Synchronization [9, 11, 12]), and was applied to deduce qualitative and quantitative performances results of adaptive algorithms.
Our aim is to claim the common idea that the CLMS is slower than the LMS, and to emphasize the existence of situations where CLMS outperforms LMS. Because the behavior of the CLMS as well as the LMS is highly dependent on the input[4], we focus on the determination of proper situations related to the choice of the input data, leading to better convergence speeds.
The main idea of this paper is first to show that the small step size hypothesis is invalid using on a theoretical analysis based on finite alphabet approach, and to demonstrate that at the optimum, the CLMS outperforms the LMS.
In section 2, we review the classical approaches used to analyze LMS and CLMS. In section 3, we present the finite alphabet approach and its power in determining performance results for both adaptive algorithms. In the last section, we focus on the evaluation and the determination of some situations where the CLMS has a better convergence speed than the LMS, by varying the parameters that impact the convergence speed.

2. CONVERGENCE ANALYSIS FOR SMALL STEP SIZE: LMS ALWAYS FASTER THAN CLMS

The considered structure is adaptive identification:

$$\begin{cases} e_k & = y_k - H_k^T X_k \\ H_{k+1} & = H_k + \mu f(X_k) e_k \end{cases} \quad (1)$$

where $X_k = [x_k, \ldots, x_{k-L+1}]^T$ is the stationary input observation vector, L is the dimension of the adaptive filter H_k, μ is a positive step size and y_k is the output signal.
If f is equal to the identity, the considered algorithm is the LMS. When $f = sign\{.\}$, we treat the case of the CLMS, where $sign\{.\}$ is the sign function.
The transient behavior of the algorithm can be described by the evolution of the deviation vector $V_{k+1} = H_{k+1} - H^{opt}$, where H^{opt} is the optimal solution of the system.
Referring to equation (1), we have:

$$V_{k+1} = (I - \mu f(X_k) X_k^T) V_k + \mu f(X_k) b_k \quad (2)$$

Note that $b_k = y_k - (H^{opt})^T X_k$, which is considered to be independent of X_k.
The performances of adaptive filters were studied under restrictive hypothesis due to the dependence of X_k and V_k. In fact, it is not easy to find an explicit relationship between $E(V_{k+1} V_{k+1}^T)$ and $E(V_k V_k^T)$, in order to describe the transient

mean square behavior of the LMS.

This is even more difficult with the CLMS since it uses the non-linear function *sign*. Therefore, the CLMS performances are usually studied and compared to the LMS, considering the following assumptions:

- continuous excitation of the algorithm,
- the input or its statistical moments are measurable,
- X_k and V_k are independent, which is only valid when adaptation is done with small values of step size μ. Generally towards zero.

The last hypothesis is very restrictive, and unrealistic, so the comparison result of LMS and CLMS determined in [1] may not be usually correct.

Under these assumptions, the mean square behavior is derived as:

$$E(V_{k+1}V_{k+1}^T) = \Delta E(V_k V_k^T) + \mu^2 E(f(X_k)f(X_k)^T(b_k)^2) \quad (3)$$

- $\Delta = (I - \mu(R_f + R_f^T) + \mu^2 S_f)$
- $S_f = E(f(X_k)X_k^T X_k f(X_k)^T)$ and $R_f = E(f(X_k)X_k^T)$

According to Equation (3), the performances results, particularly the convergence speed, depend on the quantity $\min_\mu(\delta_{max}(\mu))$, where δ_{max} is the largest eigenvalue of Δ [11].

Hence, the comparison between LMS and CLMS, will be determined, respectively, from $\min(\delta_{max}(\mu))$, which depends on both matrices R_f and S_f.

Therefore, to compare the convergence behavior of both LMS and CLMS, it was shown in [1, 3], that the CLMS pseudo-covariance matrix $R_{sign} = E(sign(X_k)X_k^T)$ is related to the covariance matrix $R = E(X_k X_k^T)$ as:

$$R_{sign} = \beta \frac{R}{\sqrt{P_x}} \quad (4)$$

where P_x is the signal's power.

For an iid and zero-mean sequence $\beta = \frac{E(\bar{x}_k^* x_k)}{[E(|x_k^2|)]^{1/2}}$.

It is proved in [1] that $0 < \beta < 1$.

The relation (4) is evaluated for small values of step size, and holds where the input signal is assumed to be zero-mean and satisfying some conditions such as iid when it is complex valued.

Based on the result (4), and noting $\eta(n) = E(|V(n)|^2)$, it was established that:

- The optimized CLMS algorithm:

$$\eta(n+1) = \eta(n) - \beta^2 \mu_{opt}^{LMS} \varepsilon(n) \quad (5)$$

- The optimized LMS algorithm:

$$\eta(n+1) = \eta(n) - \mu_{opt}^{LMS} \varepsilon(n) \quad (6)$$

where ε is the Mean Square Error.

Since $\beta < 1$, comparing the convergence speed of the mean square deviation of both algorithms (equations (5) and (6)), reveals that CLMS is slower than LMS when both of them are working at their best with step size μ_{opt}.

The value of the optimal step size is related to the critical step size μ_c defined as:

$$\mu_c = \frac{2\beta}{L\sqrt{P_x}} \quad (7)$$

A suitable choice of the optimal situation is considered where:

$$\mu_{opt} = \frac{\mu_c}{2} \quad (8)$$

Hence, we can conclude that even working at their best, using the classical approach, the comparison of the convergence speed shows that LMS is faster than CLMS.

In the next section, we present the finite alphabet approach which considers the characteristics of the input, and does not use the independence hypothesis. Thus, it allows us to determine exactly the eigenvalues of matrices that lead the system behavior. We also demonstrate, that the optimal step size μ_{opt} is larger than small step size interval, and the CLMS, at the optimum can be faster than the LMS.

3. CONVERGENCE ANALYSIS WITH QUANTIZED DATA: CLMS CAN OUTPERFORM LMS

In the context of digital transmission, the quantized input X_k belongs to a finite alphabet set $A = \{M_1, M_2, \ldots, M_N\}$, where N is the cardinality.

Since X_k is stationary, it can be modeled by a discrete-time Markov chain $\{\psi(n)\}$ with finite state space $\{1, 2, \ldots, N\}$ characterized by its probability transition matrix $P = [p_{ij}]$ such as:

$$X_k = M_{\psi(k)}$$

The finite alphabet approach consists in splitting the quantity $E(V_k V_k^T)$ into N components as follows [9]:

$$\begin{aligned} E(V_k V_k^T) &= \sum_{i=1}^{N} Q_i(k) \\ Q_j(k) &= E(V_k V_k^T 1_{\psi(k)=j}) \end{aligned} \quad (9)$$

where $1_{\psi(n)=j}$ is the indicator function.

Consequently, to establish the relationship between $E(V_k V_k^T)$ and $E(V_{k+1}V_{k+1}^T)$, we set the following recursion [9]:

$$\begin{aligned} Q_j(k+1) &= \sum_{i=1}^{N} E(V_{k+1}V_{k+1}^T 1_{\psi(k+1)=j} 1_{\psi(k)=i}) \\ &= \sum_{i=1}^{N} E((I - \mu f(X_k)X_k^T)V_k V_k^T (I - \mu f(X_k)X_k^T)^T 1_{\psi(k+1)=j} 1_{\psi(k)=i}) \\ &+ \sum_{i=1}^{N} \mu^2 E(f(X_k)(b_k)^2 f(X_k)^T 1_{\psi(k+1)=j} 1_{\psi(k)=i}) \end{aligned} \quad (10)$$

Since X_k belongs to a finite alphabet set $A = \{M_1, M_2, \ldots, M_N\}$, we can derive:

$$\begin{aligned} Q_j(k+1) &= \sum_{i=1}^{N} (I - \mu f(M_i)M_i^T) Q_j(k) (I - \mu f(M_i)M_i^T)^T p_{ij} \\ &+ \sum_{i=1}^{N} \mu^2 E((b_k)^2) p_{ij} f(M_i) f(M_i)^T \end{aligned} \quad (11)$$

If we denote $\tilde{Q}(k) = [vec(Q_1(k))^T vec(Q_2(k))^T \ldots vec(Q_N(k))^T]^T$, where vec is the operator that transforms a matrix $(n \times m)$ to a vector $(n.m)$. The linear and compact relation that governs the system (1) is:

$$\tilde{Q}(k+1) = \Gamma \tilde{Q}(k) + \tilde{Z} \quad (12)$$

Where:

$$\Gamma = (P^T \otimes I_{L^2}) diag((I_L - \mu f(M_i) M_i^T)^T \otimes (I_L - \mu f(M_i) M_i^T)) \quad (13)$$

$$\tilde{Z} = \sum_{i=1}^{N} \mu^2 E((b_k)^2) p_{ij} f(M_i) f(M_i)^T$$

$diag(.)$ is the operator allowing to construct diagonal matrix, and $\otimes$ is the Kronecker product.
The performances can be exactly determined from equation (12), more precisely from the eigenvalues of the matrix Γ as defined in (13), which depends on:

- the signal statistics: defined from the transition matrix P,
- the finite alphabet related to the different states M_i,
- the step size μ.

In fact, from matrix Γ, we can exactly determine the convergence speed of the algorithm, when working at its best conditions. Besides, according to finite alphabet approche the optimal step size μ_{opt}, allowing a faster convergence of the algorithm, is deduced from the the largest eigenvalue λ_{max} of Γ as:

$$\mu_{opt} = arg(min(\lambda_{max}(\mu))) \quad (14)$$

For the LMS algorithm, referring to equation (13), the behavior of the system (1), depends on :

$$\Gamma_{LMS} = (P^T \otimes I_{L^2}) diag((I_L - \mu M_i M_i^T)^T \otimes (I_L - \mu M_i M_i^T)) \quad (15)$$

The corresponding behavior matrix of the CLMS algorithm is:

$$\Gamma_{CLMS} = (P^T \otimes I_{L^2}) \quad (16)$$
$$diag((I_L - \mu sign(M_i) M_i^T)^T \otimes (I_L - \mu sign(M_i) M_i^T))$$

Under the small step size assumption, the conventional LMS and the CLMS comparison is based, referring to (eq(3)), on the evaluation of the eigenvalue $(min(\delta_{max}(\mu)))$ depending on the matrices R_{sign} and R as detailed in the previous section 2.
In case of quantized input data, the comparison of the two algorithms LMS and CLMS still based on the study of the largest eigenvalues, nevertheless we are interested on the parameters (λ_{max}^{LMS} and λ_{max}^{CLMS}) of the corresponding matrices (16) and (15).
This study will allow us to determine the corresponding optimal step size in order to evaluate the performances of both algorithms when working at their best, for the same conditions (alphabet, transition matrix).

4. CLIPPED LMS FASTER THAN LMS : ILLUSTRATIVE EXAMPLES

In this section, we concentrate on the determination of possibilities where CLMS is faster than LMS.
Therefore we are looking for situations satisfying the following inequality (17):

$$\frac{\min(\lambda_{max}^{LMS})}{\min(\lambda_{max}^{CLMS})} > 1 \quad (17)$$

We begin our study by a simple analytical example to mathematically prove the existence of step size range where LMS is slower than CLMS.
This is illustrated by a simulation of an adaptive identification scheme using LMS and CLMS as adaptive algorithms, and considering the matrix Γ (eq.13), relative to an appropriate alphabet set.

4.1 Analytical example

We consider a simple identification scheme with $L = 1$, an iid symmetric alphabet set $\{\pm1, \pm3\}$. The appropriate Markov chain will be characterized by the (4×4) transition matrix P with $p_{ij} = \frac{1}{4}$.
According to (15) and (16), the corresponding matrices Γ_{LMS} and Γ_{CLMS} that govern the system (1), for LMS and CLMS algorithms are respectively:

$$\Gamma_{LMS} = \frac{1}{4} \begin{pmatrix} (1-\mu)^2 & (1-\mu)^2 & (1-9\mu)^2 & (1-9\mu)^2 \\ (1-\mu)^2 & (1-\mu)^2 & (1-9\mu)^2 & (1-9\mu)^2 \\ (1-\mu)^2 & (1-\mu)^2 & (1-9\mu)^2 & (1-9\mu)^2 \\ (1-\mu)^2 & (1-\mu)^2 & (1-9\mu)^2 & (1-9\mu)^2 \end{pmatrix}$$

$$\Gamma_{CLMS} = \frac{1}{4} \begin{pmatrix} (1-\mu)^2 & (1-\mu)^2 & (1-3\mu)^2 & (1-3\mu)^2 \\ (1-\mu)^2 & (1-\mu)^2 & (1-3\mu)^2 & (1-3\mu)^2 \\ (1-\mu)^2 & (1-\mu)^2 & (1-3\mu)^2 & (1-3\mu)^2 \\ (1-\mu)^2 & (1-\mu)^2 & (1-3\mu)^2 & (1-3\mu)^2 \end{pmatrix}$$

We note that the above matrices have three null eigenvalues. Consequently, the largest eigenvalues of the two matrices are:

$$\begin{matrix} \lambda^{LMS}(\mu) = 1 - 10\mu + 41\mu^2 \\ \lambda^{CLMS}(\mu) = 1 - 4\mu + 5\mu^2 \end{matrix} \quad (18)$$

It is easy to verify that $\lambda^{CLMS} < \lambda^{LMS}$ for $\mu > \frac{1}{6}$.

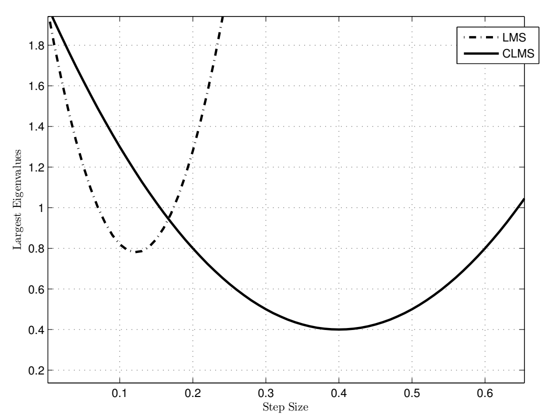


Figure 1: Largest eigenvalues versus step size for a $\{\pm1, \pm3\}$ alphabet set

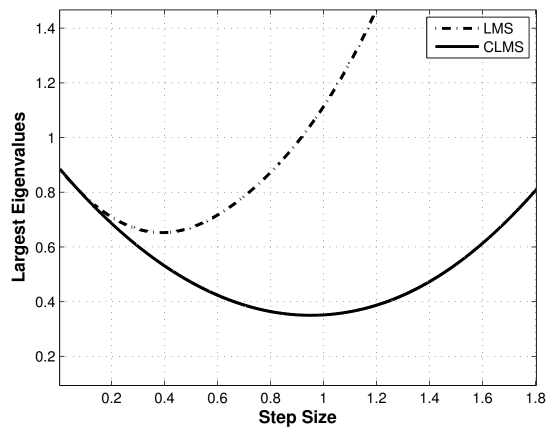
Furthermore, working at optimum we have $\min(\lambda_{max}^{LMS}) > \min(\lambda_{max}^{CLMS})$ (figure(1)), and CLMS is faster than LMS for this range of step size.

Through this simple example, it is easy to see that the common belief that LMS is faster than CLMS is only true for small step size.
We have chosen a small alphabet set, because when increasing the alphabet cardinality N and the dimension of the filter

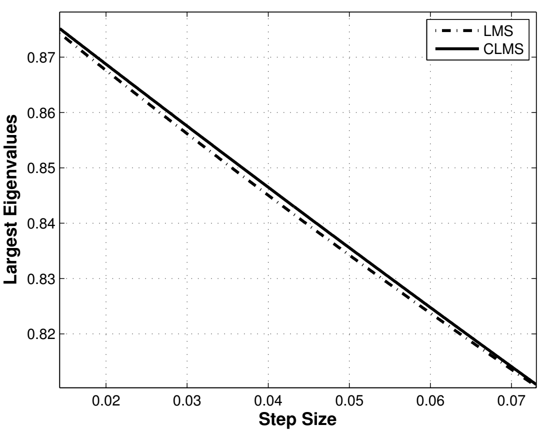
L, we deal with matrices Γ of dimension (NL^2, NL^2), and the analytical results become difficult to set. However, we can determine exactly qualitative results.

4.2 Convergence speed at optimality

Let us increase the cardinality of the input data and simulate a non correlated symmetric alphabet.
We illustrate in figure (2) the largest eigenvalues of Γ_{LMS}, and Γ_{CLMS}, when the input belongs to a finite alphabet set.



(a)Step size range

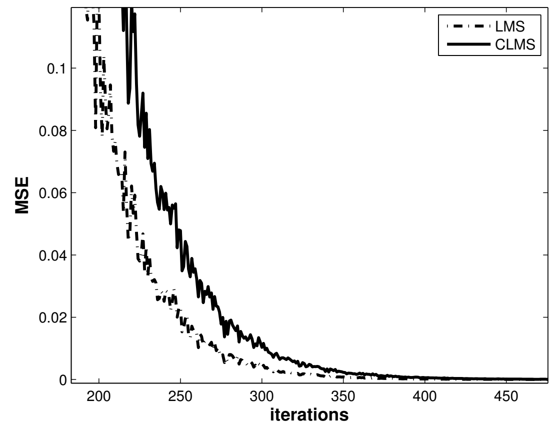


(b) Zoom for small step size

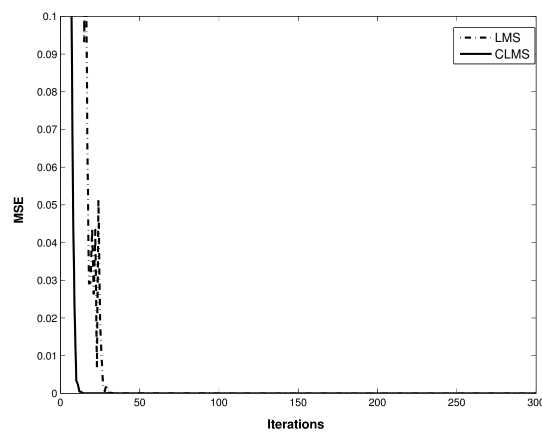
Figure 2: Largest eigenvalues versus step size for a 32 states alphabet

We observe, that for high step size varying until its critical value, λ_{max}^{CLMS} is lower than λ_{max}^{LMS}.
However, when making a zoom for small step size range (figure 2-(b)), we see that $\lambda_{max}^{LMS} < \lambda_{max}^{CLMS}$.
Even when the difference between the two eigenvalues is not very significant, we find again the results of the classical approaches.
The MSE evolution in figure (3) is simulated for an adaptive identification scheme with filter $H^{opt} = 10$, the initial condition of $H_0 = 0.01$ for both LMS and CLMS, doing 1000 Monte Carlo realizations.
We verify in figure (3) two results:

- A Classical result: Conventional LMS is faster than CLMS using small step size assumption (figure(3-(a))
- An Unexpected result: Conventional LMS is slower than CLMS as shown in (figure(3-(b)).
  The comparison was done when algorithms are working at their best using the respective optimal step size determined from the figure (2) as: $\mu_{opt}^{LMS} = 0.45$ and $\mu_{opt}^{CLMS} = 0.9$.



(a) Small step size, $\mu = 0.02$



(b) At optimality

Figure 3: Superiority of the CLMS versus LMS at optimality: MSE evolution for two ranges of step size.

For the same previous identification case, and by changing the transition matrix P of the quantized data, we draw the same conclusions at optimality.
Figure (4) emphasizes the fast convergence speed of the CLMS, compared to the LMS.

We notice that the optimal step size values are calculated from the largest eigenvalues of the matrices Γ_{LMS} and Γ_{CLMS}, and are equal to $\mu_{opt}^{LMS} = 0.07$ and $\mu_{opt}^{CLMS} = 0.85$.

5. CONCLUSION

The main purpose of this paper was to deny the common result that the conventional LMS is faster than the Clipped

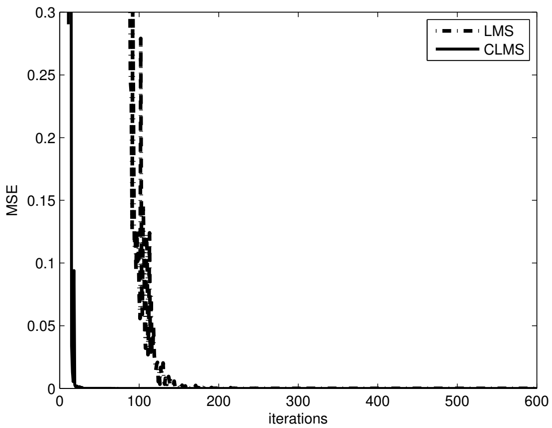


Figure 4: Second situation of fastness of the CLMS at optimality, related to the transition matrix of the input

LMS (CLMS). Through the finite alphabet approach, without using any of the classical assumptions, we have compared the convergence speed of the two algorithms, and we have determined exactly the dependence of the optimal step size on the largest eigenvalues of the matrix that governs the adaptive system.
It is worth to notice that it is possible to find other situations showing the unusual result that the CLMS is faster than the LMS. It suffices to vary the characteristics of the alphabet, namely its cardinality, its transition matrix P, which defines the input signal statistics (correlation, probability density function).

REFERENCES

[1] O. Macchi, *LMS Adaptive Processing with Applications in Transmission*. Wiley, 1994.

[2] M. Lotfizad, H.S. Yazidi, *Modified Clipped LMS Algorithm,* Eurasip Journal on Applied Signal Processing, vol. 8, pp. 1229–1234, 2005.

[3] E. Eweda, *Transient Performance Degradation of the LMS, RLS, Sign, Sigend Regressor, and Sign-Sign Algorithms with Data Correlation*, IEEE Transactions on Cicuits and Systems-II: Analog and Digital Signal Processing, Vol. 46, NO. 8, August 1999.

[4] E. Eweda, *Analysis and Design of a Sigend Regressor LMS Algorithm for Stationary and Nonstationary Adaptive Filtering with Correlated Gaussian Data*, IEEE Transactions on Circuits and Systems, Vol. 37, NO. 11, November 1990.

[5] Dirk T.M. Slock, *On the Convergence Behavior of the LMS and the Normalized LMS Algorithms*, IEEE Transactions on Signal Processing, Vol. 41, NO. 9, September 1993.

[6] Marcio.H. Costa and Jos C.M. Bermudez, *A fully Analytical Recursive Stochastic Model to the Normalized Signed Regressor LMS Algorithm*, IEEE, 2003.

[7] L.Deivasigamani, *A fast clipped-data LMS algorithm*, IEEE Transactions on Acoustics, Speech, and Signal Processing, Vol.ASSP-30, NO. 4, pp.648-649, 1982.

[8] S.Koike, *Analysis of Adaptive Filters Using Normalized Signed Regressor LMS Algorithm*, IEEE Transactions on Signal Processing, Vol. 47, NO. 10, 1999.

[9] H. Besbes, M. Jaidane-Saidane, J. Ezzine, *On Exact Convergence Results of Adaptive Filters: The finite Alphabet Case*, Eurasip Journal On Applied Signal Processing, Vol. 80, NO. 7, pp.1373-1384, 2000.

[10] N.J.Bershad, *Comments on 'Comparaison of the convergence of two algorithms for adaptive FR digital filters'*, IEEE Transactions on Acoustics, Speech, and Signal Processing, Vol.ASSP-33, NO. 6, December 1985.

[11] Y. Ben Jemaa, H. Besbes, M. Jaidane, *Does white input give better convergence speed in adaptive filtering?*, in IEEE, EUSIPCO 2000, Tampere, Finland, September 2000.

[12] M. Smaoui Kallel, Y. Ben Jemaa, M. Jaidane, *Adaptive PLL Performances Analysis for Digital Transmission Context*, IEEE, ISCCSP 04, Hammamet, Tunisia, March 2004.