DOUBLE DIRECTION ADAPTATION FOR NOISE REDUCTION IN PRE-WHITENED LMS-TYPE ALGORITHMS

Hichem Besbes, Sofia Ben Jebara

Research unit TECHTRA
Ecole Supérieure des Communications de Tunis
Route de Raoued 3.5 Km, Cité El Ghazala, 2083, Ariana, Tunisia
hichem.besbes@supcom.rnu.tn, sofia.benjebara@supcom.rnu.tn

ABSTRACT

The LMS algorithm suffers from its slow rate of convergence, especially for high correlated input signal. The input pre-whitening based algorithms provide better convergence rate with the price of noise enhancement. To mitigate this drawback, we present in this paper a technique, which consists on exciting the adaptive filter at both the input signal direction and the pre-whitened input direction. Hence, two different step sizes are used, they permit to improve convergence rate without enhancing the noise during steady state. A theoretical analysis of the steady state performance is presented. Simulation results are also presented to support the analysis and to compare the proposed algorithm with classical ones.

1. INTRODUCTION

Several adaptation algorithms are developed in order to identify unknown systems. The Least Mean Square (LMS) algorithm is the commonly used one [1]. Although its computational requirements are low, it suffers from slow rate of convergence, especially for colored input signals. To overcome this drawback, several approaches based on input decorrelating were proposed [1][2]. The main idea is to pre-whiten or decorrelate the excitation before passing it to the adaptive algorithm. This is possible by using decorrelation filters whose outputs are predictor errors. Various possibilities of placing decorrelation filters within the identification system are possible (see for example [3] for global overview on main solutions in acoustic echo cancellation field).

In this paper, the proposed solution is based on algorithm adaptation which operates at two decorrelated directions: the direction of the pre-whitened input signal, in order to fasten convergence and the direction of the input signal, in order to reduce noise enhancement generated by output filtering. Furthermore, the adaptation process is applied one time each two iterations. It leads to unwanted bursts reduction, appearing commonly in classical approaches using pre-whitening techniques.

The paper is organized as follows. In section 2, we present some backgrounds on classical techniques using input decorrelation approach. We will focus the study on their advantages and limitations. In section 3, we firstly develop the main idea characterizing the proposed algorithm. Next, we present the Double Direction Pre-whitened LMS algorithm (DDP-LMS). In section 4, we present a mathematical formulation for the steady state analysis. Section 5 is devoted to a comparison between the proposed algorithm and classical ones such as the LMS and two variants of the filtered-XLMS algoritms. Finally, some concluding remarks are provided.

2. BACKGROUNDS ON FILTERED X-LMS ALGORITHMS

2.1. Filtered X-LMS algorithms

In this paper, we consider the identification problem of a linear system characterized by the following input/output relationship:

$$y(k) = F^T X(k) + n(k), \tag{1}$$

where F is the system impulse response of length L, $X(k) = [x(k), \cdots, x(k-L+1)]^T$ is the observation vector of the input signal, and $n(k)$ is an additive white Gaussian noise.

The adaptive filter $H(k)$ is used to identify the system impulse response F. It is well known that the best convergence rate of an adaptive filter is obtained when it is excited by a quasi-white input signal. Hence, diferent algorithms use a predictor, $P(k) = [p_1(k), p_2(k), .., P_N(k)]^T$, of length N to decorrelate the input. The output of the pre-whitener is given by:

$$x^f(k) = x(k) - \sum_{i=1}^{N} p_i(k) x(k-i). \tag{2}$$

In this section, we are interested by the algorithms resumed in Table 1.

Table 1. Algorithm descriptions.

Notation	Adaptation expression
LMS	$H(k+1) = H(k) + \mu e(k) X(k)$
PI-XLMS	$H(k+1) = H(k) + \mu e(k) X^f(k)$
PIFE-XLMS	$H(k+1) = H(k) + \mu e^f(k) X^f(k)$

The $PI-XLMS$ (Pre-whitened Input XLMS) algorithm is the most popular one, due to its robustness and simplicity of implementation [1]. It is proposed in order to achieve better convergence rate than the LMS. Another approach, where the adaptive filter is driven by both the pre-whitened input and the filtered error is called Pre-whitened Input/Filtered Error XLMS algorithm (PIFE-XLMS) [4]. The filtered error is obtained using the same pre-whitening filter:

$$e^f(k) = e(k) - \sum_{i=1}^{N} p_i(k) e(k-i). \tag{3}$$

2.2. Performances

The deviation vector is the common tool to evaluate adaptive algorithm behavior [1][2]. It is defined as the difference between the adaptive filter and the unknown one:

$$V(k) = H(k) - F. \tag{4}$$

In case of LMS algorithm, it is well known that its convergence rate depends on the eigenvalue spread of the input correlation matrix $R_x = E\left\{X(k)X(k)^T\right\}$. Faster convergence is obtained for white inputs [1][2].

In order to improve the convergence rate, PI-XLMS was introduced. It is also well known that its convergence rate depends on matrix $R'_x = E\left\{X^f(k)X(k)^T\right\}$ [1]. It is a lowercase triangular matrix whose eigenvalues are identical, leading to an accelerated convergence rate when compared to that of LMS.

An alternative solution to accelerate the convergence rate is to use PIFE-XLMS. Its deviation vector obeys the following recursion

$$\begin{aligned} V_{PIFE}(k+1) = \; & \left[I - \mu X^f(k)X^f(k)^T\right] V_{PIFE}(k) \\ & +\mu n^f(k)X^f(k) - \mu^2\Delta(k), \end{aligned} \tag{5}$$

where $\mu^2\Delta(k)$ is given by:

$$\mu^2\Delta(k) = \mu \sum_{i=1}^{N} p_i \left[H(k) - H(k-i)\right]^T X(k-i)X^f(k). \tag{6}$$

This recursion permits the following interpretations:

• For small step sizes, the term on $\mu^2\Delta(k)$ can be neglected. We obtain the same performances as the LMS algorithm excited with a quasi-white input $x^f(k)$ and whose additive noise is $n^f(k)$. The convergence rate is then the best one.

• During steady state, we notice additive noise amplification. In fact, the filtered noise power is greater than additive noise power ($P_{n^f} = P_n \sum_{i=1}^{N} p_i^2 > P_n$. Such noise amplification constitutes the main drawback of the PIFE-XLMS algorithm.

• For large step sizes, the term on $\mu^2\Delta(k)$ has an effect on algorithm performances. It depends on past errors which render the algorithm sensitive for error variations. In fact, in case of important errors variations, local bursts can appear, degrading locally the performances.

2.3. Illustration

To illustrate the last interpretation, we carry the following simulation. We consider an input signal of power $P_x = 10$ modeled by AR(1) model $x(k) = \rho x(k-1) + g(k)$, where $g(k)$ is a white noise and $\rho = 0.95$. The system to be identified is characterized by an impulse response $F = [1;\, 0;\, 10;\, -6;\, -1;\, 4;\, 0.1;\, 5;\, -2;\, -0.1]^T$ ($L = 10$), and an additive noise of power $P_n = 0.1$. The tested algorithms are the PI-XLMS, the PIFE-XLMS, and the LMS excited with white input of power $P_{x^f} = \left(1-\rho^2\right)P_x$. Two different additive noises are considered. The first one is AWGN $n(k)$ of power P_n. The second one is $n^f(k) = n(k) - \rho n(k-1)$ of power $P_{n^f} = \left(1+\rho^2\right)$.

Fig. 1 illustrates the evolution of the Mean Square Deviation $MSD(k) \stackrel{\triangle}{=} E\left\{V(k)^T V(k)\right\}$ versus iteration number for $\mu = 0.005$. It can be observed that PIFE-XLMS has the same transient state as LMS with white input and it has better convergence rate than PI-XLMS. During steady state, we notice degradation. In fact, some local bursts are observed and the error power is close to that of LMS with additive noise $n^f(k)$.

3. DOUBLE DIRECTION PREWHITED LMS ALGORITHM

3.1. Motivations and preliminary solutions

Our purpose is to propose a solution to improve PIFE-XLMS by taking advantage of convergence improvement with an adequate correction to mitigate the noise enhancement and the algorithm sensitivity to errors. For such purpose, we propose to cancel the term $\mu^2\Delta$ on 5. It is possible by setting $H(k) = H(k-i)\ \forall\ i = 1..N$. This operation is equivalent to adaptation stop during N iterations. More precisely, adaptation is carried at iterations whose index is multiple of $N+1$ and stopped otherwise.

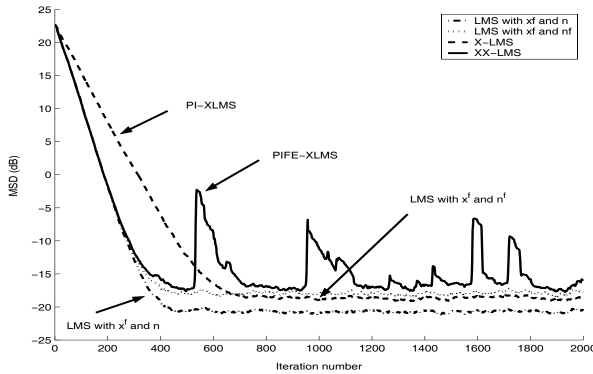

Fig. 1. Illustration of PIFE-XLMS negative aspects for large step size.

Furthermore, we propose to limit adaptation stop to one iteration over two, and use a first order predictor. It is justified by two critical points [5]:

• During transient state, the convergence time is N times the one of white input. The larger the N, the slower the convergence rate.

• During steady state, the level of error is amplified. The improvement in term of convergence rate will be counterbalanced by error amplification.

3.2. Adaptation at two directions

The proposed idea to avoid error amplification in illustrated in Fig. 2 and is described as follows:

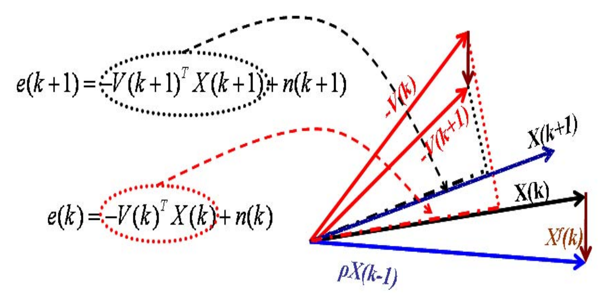

Fig. 2. Illustration

• The transient behavior of an adaptive can be analyzed through the evolution of the error signal $e(k)$, which can be expressed by:

$$e(k) = -V(k)^T X(k) + n(k). \tag{7}$$

When ($V(k)^T X(k) \approx 0$), the error is minimized, this means that the deviation vector is orthogonal to the input observation vector, and we have excited the algorithm $H(k)$ in the direction of $X(k)$.

• For high correlated inputs characterized by $\frac{|X(k+1)^T X(k)|}{\|X(k+1\| \, \|X(k)\|}$ close to 1, and by using an optimal predictor (leading to white $x^f(k)$ signal), the observation vector of the pre-whitened input is orthogonal to $X(k)$ and quasi-orthogonal to $X(k+1)$.

• Using PIFE approach, we adapt at $X^f(k)$ direction which permits to attenuate the deviation vector ($\|V(k+1)\| \leq \|V(k)\|$). However, according to Fig. 2, we observe that the projection of $V(k+1)$ on

$X(k+1)$, $V(k+1)^T X(k+1)$, didn't decrease so much.
By exploiting the property of $e(k)$ in (7), we can deduce that by updating the algorithm in the $X(k)$'s direction, we will get $V(k)^T X(k)$ close to zero. Hence we reduce the amplitude of the error signal and we improve the convergence rate of the algorithm.

3.3. Double Direction Pre-whitened LMS algorithm

The main idea of the proposed algorithm is to excite the algorithm on two directions: the direction of the pre-whitened signal $X^f(k)$, in order to fasten the convergence of the algorithm, and the direction the input signal $X(k)$ in order to reduce the mean square error during the transient phase. The proposed algorithm expressions are re-formulated as follows:

$$\left\{\begin{array}{rcl} H(2k+1) &=& H(2k) \\ H(2k+2) &=& H(2k)+\mu_f e^f(2k+1)X^f(2k+1) \\ && +\mu_x e(2k)X(2k). \end{array}\right. \quad (8)$$

4. STEADY STATE ANALYSIS

4.1. General formulation

The behavior of the proposed DDP-LMS is described by the deviation vector recursion

$$\begin{array}{rl} V(2k+2) = & [I-\mu_f M_f(2k+1)-\mu_x M_x(2k)]\,V(2k), \\ & +\mu_f X^f(2k+1)n^f(2k+1)+\mu_x X(2k)n(2k), \end{array} \quad (9)$$

where $M_x(2k)=X(2k)X(2k)^T$ and $M_f(2k+1)=X^f(2k+1)X^f(2k+1)^T$. The mean square behavior is described through the evolution of the auto-correlation of the deviation vector defined by

$$R_V(k) \stackrel{\triangle}{=} E\left\{V(k)V(k)^T\right\}. \quad (10)$$

Using equation (9), and the well known independence assumption, it is easy to show that $R_V(k)$ satisfies the following recursion:

$$\begin{array}{rl} R_V(2k+2) = & R_V(2k)+\mu_x^2\Gamma_{x,x}++\mu_f^2\Gamma_{f,f}+\mu_x\mu_f\Gamma_{x,f} \\ & -\mu_f\left[E\{M_f(2k+1)\}R_V(2k)+R_V(2k)E\{M_f(2k+1)\}\right] \\ & -\mu_x\left[E\{M_x(2k)\}R_V(2k)+R_V(2k)E\{M_x(2k)\}\right] \\ & +\mu_f^2 P_{nf}E\{M_f(2k+1)\}+\mu_x^2 P_n E\{M_x(2k)\} \\ & +\mu_f\mu_x E\left\{n(2k)n^f(2k+1)\right\}\times \\ & \left[E\left\{X^f(2k+1)X(2k)^T\right\}+E\left\{X(2k)X^f(2k+1)^T\right\}\right], \end{array} \quad (11)$$

where

$$\left\{\begin{array}{rl} \Gamma_{x,x} = & E\{M_x(2k)R_V(2k)M_x(2k)\} \\ \Gamma_{f,f} = & E\{M_f(2k+1)R_V(2k)M_f(2k+1)\} \\ \Gamma_{f,x} = & E\{M_x(2k)R_V(2k)M_f(2k+1)\} \\ & +E\{M_{xf}(2k+1)R_V(2k)M_x(2k)\}. \end{array}\right. \quad (12)$$

In theoretical point of view, (11) permits to analyze all mean square performances in both transient and steady state. However, in practical point of view, (11) is hard to solve due to its complicate structure. In order to overcome these difficulties, we used tensorial algebra properties [5]. The obtained expressions permits exact performances analysis. However, they are presented in compact form which renders their interpretation difficult. In this paper, we propose to approximate the steady state performance for small step sizes. This approach will proove in concrete equations the importance of adaptation at two directions and other properties, reinforcing the validity of the proposed algorithm.

4.2. Small step sizes analysis

Without loss of generality, we consider the special case of AR(1) input signal with correlation coefficient ρ . We assume that the optimal predictor $p(k)=\rho$ is used. In this case, $x^f(k)$ is white of power $P_{xf}=(1-\rho^2)P_x$ and the filtered noise power is $P_{nf}=(1+\rho^2)P_n$.

We focus our analysis on small step sizes. In such case, the terms $\mu_x^2\Gamma_{x,x}$, $\mu_f^2\Gamma_{f,f}$ and $\mu_x\mu_f\Gamma_{x,f}$ in (11) are neglected. Under the convergence conditions ($R_V(2k+2)=R_V(2k)$ for $k\to\infty$), we can show easily that the auto-correlation matrix R_V satisfies:

$$\begin{array}{l} \mu_x\left[R_V R_x+R_x R_V\right]+2\mu_f P_{xf}R_V = \\ \mu_f^2 P_{nf}P_{xf}I+\mu_x^2 P_n R_x-\mu_x\mu_f P_n\dfrac{P_{xf}}{P_x}(R_x-P_x I) \end{array} \quad (13)$$

In order to solve (13), we denote Λ, a diagonal matrix whose elements λ_i are the eigenvalues of the input auto-correlation matrix $R_x=Q\Lambda Q^{-1}$. By the same way, we denote $R_V=Q\Phi Q^{-1}$. Equation (13) can be rewritten as follows:

$$\begin{array}{l} \mu_x(\Phi\Lambda+\Lambda\Phi)+2\mu_f P_{xf}\Phi = \\ \mu_f^2 P_{nf}P_{xf}I+\mu_x^2 P_n\Lambda-\mu_x\mu_f P_n\dfrac{P_{xf}}{P_x}(\Lambda-P_x I) \end{array} \quad (14)$$

The first step consists on calculating the diagonal elements of the matrix Φ. Next, the excess mean square error ($EMSE$) approximated by:

$$EMSE \approx trace(R_x R_v), \quad (15)$$

can be calculated by taking the trace of (14). It is given by:

$$EMSE(\nu_x,\nu_f) = \begin{array}{l} \dfrac{P_n\nu_x}{2}+\dfrac{P_n\nu_f}{2}\dfrac{1+\rho^2}{1-\rho^2} \\ -P_n\frac{\nu_x\nu_f}{1-\rho^2}\left\{\frac{1}{L}\sum_{i=1}^{L}\frac{c_i^2}{\nu_f+c_i\nu_x}\right\} \end{array}. \quad (16)$$

Where $\nu_x=\mu_x LP_x$, $\nu_f=\mu_f LP_{xf}$ and $c_i=\dfrac{\lambda_i}{P_x}$ are the normalized eigenvalues.

Two special cases can be derived. The first one (resp. second one) is obtained by adaptation at pre-whitened input direction (resp. input direction). It corresponds to the performance of the PIFE-XLMS (resp. LMS) algorithm.

$$\begin{array}{rl} EMSE(0,\nu_f) & = \frac{P_n\nu_f}{2}\frac{1+\rho^2}{1-\rho^2} \\ EMSE(\nu_x,0) & = \dfrac{P_n\nu_x}{2}. \end{array} \quad (17)$$

4.3. Bounding $EMSE$

In order to point out the importance of exciting the algorithm in the direction of the input signal, we propose to determine an upper bound of the $EMSE$. For such purpose, we consider the properties of the function $h(c_i)=\dfrac{c_i^2}{\nu_f+c_i\nu_x}$. Since ν_f and ν_x are positives, the function $h(.)$ is convex in $[0,+\infty[$

$$\sum_{i=1}^{L}\frac{1}{L}h(c_i) \geq h\left(\sum_{i=1}^{L}\frac{1}{L}c_i\right), \quad (18)$$

Since $c_i=\dfrac{\lambda_i}{P_x}$, it is easy to show that: $\sum_{i=1}^{L}\frac{1}{L}c_i=1$. Hence, the lower bound is given by:

$$h\left(\sum_{i=1}^{L}\frac{1}{L}c_i\right)=\frac{1}{\nu_f+\nu_x}. \quad (19)$$

The $EMSE$ can be bounded as follows:

$$\begin{aligned} EMSE(\nu_x,\nu_f) \leq & \quad B_{EMSE}(\nu_x,\nu_f) \\ B_{EMSE}(\nu_x,\nu_f) = & \quad \frac{P_n\nu_f}{2}\frac{1+\rho^2}{1-\rho^2} \\ & -\frac{P_n\nu_x}{2(1-\rho^2)}\left\{\frac{(1+\rho^2)\nu_f-(1-\rho^2)\nu_x}{\nu_f+\nu_x}\right\} \end{aligned} \quad (20)$$

From (20), we can show that it exists an optimal step size ν_x^{opt}, which minimizes the upper bound $B_{EMSE}(\nu_x,\nu_f)$. It is given by

$$\nu_x^{opt} = \left(-1+\sqrt{\frac{2}{1-\rho^2}}\right)\nu_f. \quad (21)$$

The minimum upper bound $B_{EMSE}(\nu_x^{opt},\nu_f)$ is given by:

$$B_{EMSE}(\nu_x^{opt},\nu_f) = \frac{P_n\nu_f}{2}\frac{1+\rho^2}{1-\rho^2} - \frac{P_n\nu_f}{2}\frac{1+\rho^2}{1-\rho^2}\frac{\sqrt{2}-\sqrt{1-\rho^2}}{\sqrt{2}-\sqrt{1+\rho^2}}. \quad (22)$$

Furthermore, we can show that this last term is also bounded by the $EMSE$ obtained by adaptation at $X_f(k)$ direction (17):

$$B_{EMSE}(\nu_x^{opt},\nu_f) \leq EMSE(0,\nu_f), \quad (23)$$

This means that for every $\nu_f \neq 0$, it exist ν_x^{opt} which gives better steady state than the one obtained with adaptation at filtered input direction only:

$$EMSE(\nu_x^{opt},\nu_f) \leq B_{EMSE}(\nu_x^{opt},\nu_f) \leq EMSE(0,\nu_f). \quad (24)$$

Fig. 3 illustrates such interpretation for a two-tap system described by $F = [1; -0.95]$ and an AR(1) input with $\rho = 0.95$. The step size μ_x is adjusted according to (21). We effectively show both better steady state and better convergence rate when compared to the same algorithm with adaptation at the direction of $X^f(k)$ only.

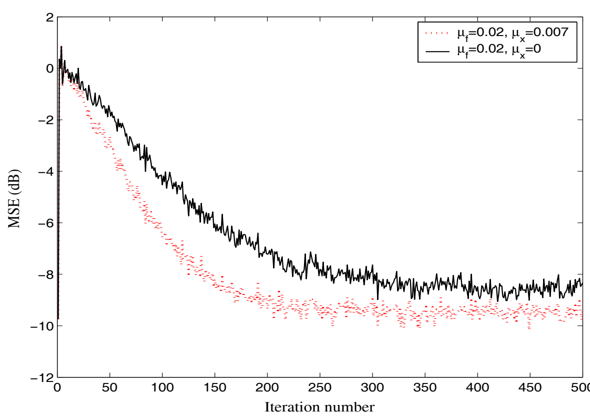

Fig. 3. Illustration of importance of adaptation at two directions.

5. COMPARISON WITH CLASSICAL ALGORITHMS

To point out the effectiveness of our algorithm, we consider an input signal modeled by an auto regressif model, AR(5), with parameters $[0.9; -0.7; 0.3; -0.6; 0.2]$. The impulse response of the system to be identified is $F = [1; 0; 10; -6; -1; 4; 0.1; 5; -2; -0.1]^T$ ($L = 10$). The additive noise has a power of $P_n = 0.1$. The tested algorithms are the DDP-LMS, the PI-XLMS, the PIFE-XLMS, and the LMS. To compare the rate of convergence, we choose the case

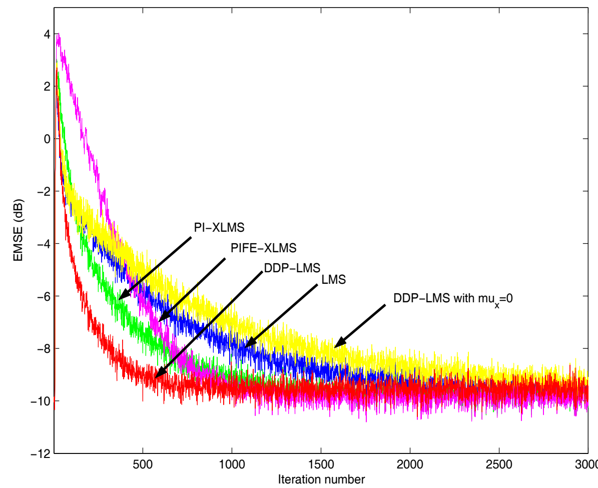

Fig. 4. Convergence rate comparison.

where all algorithms achieve the same steady state. Fig. 4 shows the evolution of the $EMSE$ in the yet mentioned simulation conditions. We note that the performance of the proposed algorithm are better than those of the LMS, PI-XLMS and PIFE-XLMS. We also notice the rate convergence improvement when we adapt at the two directions.

6. CONCLUSION

In this paper, we have presented a new algorithm tailored for high correlated input signals. It is based on combination of LMS and filtered LMS algorithms. The adaptation process is carried out on two directions, in the direction of the input and in the direction of the pre-whitened input. An analytical study of steady state performances is presented, it justifies that the proposed structure have better steady state performance when adaptation is carried at two directions. Furthermore, simulation results are presented to support the proposed algorithm and to compare it to classical filtered XLMS algorithms.

7. REFERENCES

[1] B. Widrow and S. D. Stearns, "Adaptive Signal Processing," *Prentice-Hall*, 1985.

[2] S. Haykin, "Adaptive Filter Theory," *3rd ed., Englewood Clifs, NJ:Prentice-Hall*, 1996.

[3] C. Breining, P. Dreiseitel, E. Hansler, A. Mader, B. Nitsch, H. Puder, T. Schertler, G. Schmidt and J. Tilp, "Acoustic echo control. An application of very-high-order adaptive filters," *IEEE Signal Processing Magazine*, vol. 16, no. 4, pp. 42-69, July 1999.

[4] M. Mboup, M. Bonnet and N. Bershad, "LMS coupled adaptive prediction and system identification: a statisical model and transient mean analysis," *IEEE Trans. on Signal Processing*, vol. 42, No 10, pp. 2607-2615, Oct. 1994.

[5] H. Besbes and S. Ben Jebara, "A solution to reduce noise enhancement in pre-whitened LMS-type algorithms : The double direction adaptation," on *Proceedings Int. Symposium on Control, Communications and Signal Processing ISCCSP*, IEEE, 2004.