A VARIABLE STEP SIZE PRE-WHITENED SIGN ALGORITHM WITH QUANTIZED NORMALIZING FACTOR. APPLICATION TO ACOUSTIC ECHO CANCELLATION

Hichem Besbes and Sofia Ben Jebara

Département de Mathématiques Appliquées, Signal et Communications,
Ecole Supérieure des Communications de Tunis
Cité Technologique des Communications de Tunis,
Route de Raoued 3.5 Km, Cité El Ghazala, 2083, Ariana, Tunisia
phone: +216 71 857 000, fax: +216 71 856 829,
email:hichem.besbes@supcom.rnu.tn, sofia.benjebara@supcom.rnu.tn

ABSTRACT

In this paper, a variable step size pre-whitened sign algorithm with a quantized normalizing factor is proposed and applied for acoustic echo cancellation. The normalizing factor, introduced in adaptive algorithms, makes filter insensitive to variations in filter input power. The Quantization of this factor to the nearest power of two, leads to a low implementation cost. Furthermore, the concept of input pre-whitening is used to drive the adaptive filter, yielding to a faster convergence rate. To make the acoustic echo canceller robust, a new way of varying the step size is proposed. The main idea of varying the step size is very simple and depends on the state of the algorithm: transient state, steady state, or “double talk” state. It was finding that it improves significantly the ability of the algorithm to tackle the problem of double talk presence, and contributes in increasing the convergence rate by maintaining a good steady state performance. Simulation results are presented to support the proposed algorithm.

1. INTRODUCTION

The acoustic echo is due to the feedback of the far-end speaker’s voice through the loudspeaker-microphone path. Teleconferencing systems, hands-free telecommunication terminals, and voice over IP systems employ acoustic echo cancellers to reduce such echo. These systems are based on adaptive filtering principle. Since the echo path is characterized by a long impulse response, fast convergence and robust algorithms are needed. This in turn leads to sophisticated systems and high computational complexity (see [1] for example).

Since the input speech signal to an acoustic echo canceller is non-stationary in nature, normalized versions of adaptive algorithms are used in order to track signal variations. However, supplementary multiplication and division operations are required to accomplish this normalization. In this paper, we propose to normalize the sign algorithm. In this case, the normalization factor is computed only by using addition. Furthermore, we quantize the normalization factor to the nearest power of two. The division operation is then assimilated to a simple shifting operation.

The signed adaptive algorithms are mainly proposed in order to reduce complexity. However, they suffer from slow convergence. In this paper, we aim improving sign algorithms performance by using pre-whitening concept. In fact, as for stochastic gradient adaptive algorithms, input correlation degrades sign algorithm performances. Hence, pre-whitened sign algorithm increases the convergence rate [2].

The pre-whitener is also driven by an adaptive algorithm because of the non stationarity of the input speech signal. It is common to use the Normalized Least Mean Square Algorithm (NLMS) [3]. This in turn leads to computational complexity increase. We propose to overcome this drawback by using sign-based pre-whitener.

To ensure a good trade-off between fast convergence rate and stability during double talk, a variable step size is used. Basically we improve the idea of the Dual Sign Algorithm (DSA)[4] by using three step sizes [5], depending on the state of the algorithm. We define three states: transient state, steady state, and “double talk” state. The switching rules between step sizes depend on the previous state and employ some kind of hysteresis to ensure algorithm stability and effectiveness.

2. MOTIVATION

2.1 Background

The classical method used for acoustic echo cancellation is based on adaptive identification of the impulse response F of the echo path. The echo canceller generates an echo replica $\hat{y}(k)$ by filtering the speech input $x(k)$ by an adaptive filter $H(k)$. This replica is subtracted from the microphone signal $y(k)$ yielding to the near-end signal estimation $e(k)$, which is used to control adaptively the filter $H(k)$. The echo canceller can be resumed in the following equation:

$$\left\{ \begin{array}{rl} y(k) = & F^T X(k) + n(k) \\ \hat{y}(k) = & H(k)^T X(k) \\ e(k) = & y(k) - \hat{y}(k) \\ H(k+1) = & H(k) + \mu(k)\,\phi\,(e(k))\,\psi\,(X(k)) \end{array} \right. \tag{1}$$

Where $n(k)$ represents the sum of the noise and the near-end speech, $X(k) = [x(k), x(k-1), \cdots x(k-L+1)]^T$ is the observation vector of the input signal, L is the system impulse response length, $\mu(k)$ is a step size, $\phi(e(k))$ and $\psi(X(k)$ are two functions characterizing the algorithm.

In this paper, we focus more on the family of normalized algorithms [6]. In this case, the step size is given by:

$$\mu(k) = \frac{\mu}{\mathcal{N}(k) + \beta}, \tag{2}$$

where μ is a positive step size, β is a regularization parameter, and $\mathcal{N}(k)$ is a normalization factor.

From equations 1 and 2, we can derive different well known algorithms:

1. if $\phi(e(k)) = x(k)$, $\psi(X(k)) = X(k)$ and $\mathscr{N}(k) = X(k)^T X(k)$, the algorithm becomes the well known Normalized LMS algorithm.
2. if $\phi(e(k)) = e(k)$, $\psi(X(k)) = sign(X(k))$ and $\mathscr{N}(k) = \sum_{i=0}^{L-1} |x(k-i)|$, the algorithm becomes the Nagumo and Noda algorithm (NNA)[7].
3. if $\phi(e(k)) = sign(e(k))$, $\psi(X(k)) = X(k)$ and $\mathscr{N}(k) = \sum_{i=0}^{L-1} |x(k-i)|$, the algorithm becomes the normalized sign algorithm [8].

For highly correlated input signal, adaptive filters have a slow rate of convergence. In the next section we present our motivation to use the input pre-whitening type algorithm.

2.2 Pre-whitened Input type algorithms

It is well known that stochastic gradient adaptive filters perform best when operating on uncorrelated input signals [3]. In a previous work, this result was generalized for dual sign algorithm [2]. In this section, we point out the effect of input correlation on the rate of convergence of sign algorithm. We present in Figure 1 the evolution of the Mean Square Deviation $MSD(k)$ versus iteration number :

$$MSD(k) \stackrel{\triangle}{=} (H(k)-F)^T (H(k)-F). \tag{3}$$

We considered the following case: the input is a first order autoregressive process $x(k) = \rho x(k-1) + g(k)$, where $g(k)$ is a Gaussian white noise. We have chosen three values of ρ, namely, $\rho = 0$, $\rho = 0.5$ and $\rho = 0.9$. The system impulse response is an acoustic response of a visioconference room truncated to $L = 512$ (for a sampling frequency of 16 Khz). The additive noise is of power $P_n = 0.1$. Figure 1 shows that, for the same steady state, the convergence rate degrades when the input correlation increases considerably. However, the convergence rate for weakly correlated input($\rho = 0.5$) are close to that of white input.

We may conclude that signal pre-whitening approach can then be applied in order to enhance the convergence rate of highly correlated input such as speech, and this can be done without the need of employing a high order predictor. In fact, low correlated inputs permits quasi-equivalent quality than white inputs.

For each kind of sign algorithm (sign, sign-sign, normalized sign,...), different solutions, inspired from LMS based algorithms, are possible : pre-whitening only input, pre-whitening input and filtering the error using the same pre-whitener,...

In a previous work, we investigate the pre-whitened sign concept. We validate it by both theoretical and simulation approaches for Dual Sign Algorithm (DSA) when only the pre-whitened input drives the adaptive algorithm [2]. In this work, we propose to develop a variable step size filtered sign algorithm tailored for acoustic echo cancellation. One main advantage is to maintain low the computational complexity. More precisely, both adaptive filter used for system identification and adaptive predictor used for input pre-whitening are based on sign algorithm.

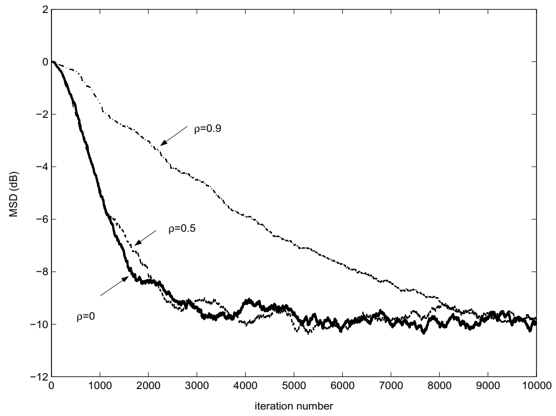

Figure 1: Sign algorithm performance with white and correlated inputs.

3. THE QUANTIZED NORMALIZING FACTOR FOR NORMALIZED ADAPTIVE ALGORITHMS

3.1 Basic idea

Since the normalizing factor introduces at least one supplementary division, we propose to reduce this complexity by quantizing this factor to the nearest power of two. The quantity $\mathscr{N}(k)+\beta$ in the expression of $\mu(k)$ given in the equation 2, can be replaced by:

$$Q(\mathscr{N}(k)+\beta) = 2^b = 2^{Int[\log_2(\mathscr{N}(k)+\beta)]}, \tag{4}$$

where Int is the operator which rounds to the nearest integer.

This operation is very cheap in terms of hardware implementation, and implies also the use of low bit width resolution to compute the normalization factor $\mathscr{N}(k)$, which reduces the cost of multiplication, if any multiplication is used in $\mathscr{N}(k)$.

To illustrate the effect of quantizing the normalizing factor on the performance of the NLMS and the NNA, we have considered the following case: the input is a second order autoregressive process $x(k) = 1.4x(k-1) - 0.45x(k-2) + g(k)$, where $g(k)$ is a Gaussian white noise. The system impulse response has a length $L = 64$. The additive noise is of power $P_n = 0.001$.

Figure 2 depicts the evolution of the MSD versus iteration number for the NLMS, NNA and their quantizing normalizing factor version, for $\mu = 2^{-1}$, $\beta = 2^{-6}$. It shows that, quantizing the normalizing factor does not affect considerably the performance of the algorithm.

3.2 The Quantized Normalizing factor Sign Algorithm (QNSA) applied for input pre-whitening

To obtain a system operating with quasi-white inputs, we can use an input pre-whitener. Its output $x^f(k)$ is described as follows :

$$x^f(k) = x(k) - P(k)^T \tilde{X}(k-1), \tag{5}$$

where $P(k)$ is the adaptive predictor of length L_P and $\tilde{X}(k-1) = [x(k-1), x(k-2), ..., x(k-L_P)]^T$ is the input observation vector.

Usually, an adaptive predictor is used with speech signals, it is classically driven by Normalized Least Mean

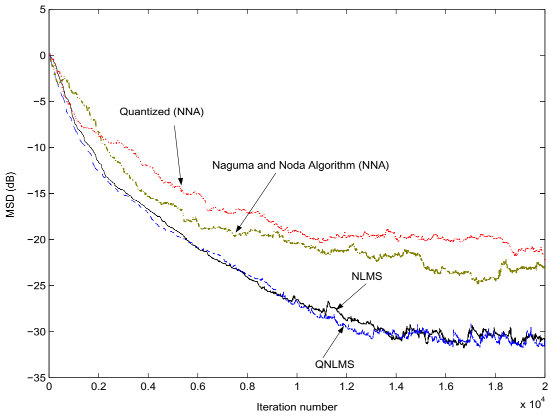

Figure 2: Effect of Quantizing the Normalizing factor on the performance of the adaptive algorithm

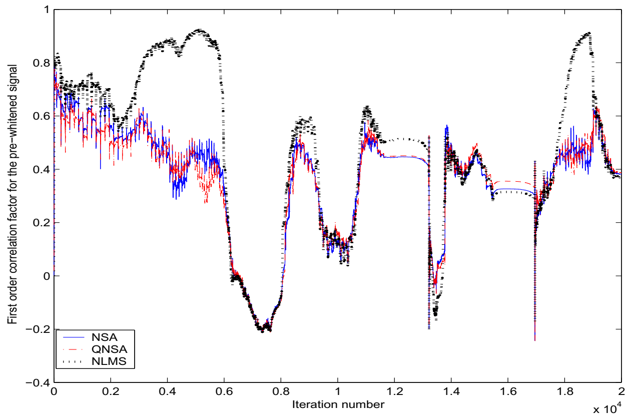

Figure 3: Evolution of the first order correlation factor of the pre-whitened signal $x^f(k)$ using NSA, QNSA and NLMS algorithms.

Square Algorithm (NLMS). In this paper, we'll investigate the use of normalized sign-based adaptive predictor. The proposed predictor is driven by the Quantized Factor Normalized Sign Algorithm (QNSA), and it is described as follows:

$$P(k+1) = P(k) + \frac{\mu_P \operatorname{sign}\left(x^f(k)\right)}{Q\left(\sum_{i=1}^{L_p} |x(k-i)| + \beta\right)} \tilde{X}(k-1), \quad (6)$$

where μ_P is the pre-whitener step size.

In order to validate such proposed idea, we have applied the adaptive predictor when the input is a speech signal sampled at 16kHz. The measure the performance of the predictor is evaluated through the first order correlation factor of the pre-whitened signal $x^f(k)$. This correlation factor is estimated as follows :

$$\begin{cases} \rho_1(k) = \dfrac{\widehat{C}_{x^f}(k)}{\widehat{P}_{x^f}(k)} \\ \widehat{C}_{x^f}(k) = \alpha \widehat{C}_{x^f}(k-1) + (1-\alpha) x^f(k) x^f(k-1) \\ \widehat{P}_{x^f}(k) = \alpha \widehat{P}_{x^f}(k-1) + (1-\alpha) x^f(k)^2 \end{cases}, \quad (7)$$

where α is the forgetting factor chosen equal to $\alpha = 0.995$.

In figure 3 we report the evolution of $\rho_1(k)$, when the predictor is driven by the NSA, QNSA and the NLMS algorithms. The step size is chosen to be the same for the three algorithms and it is equal to $\mu_P = 2^{-2}$. The regularization parameter is chosen equal to $\beta = 1$. From this figure, we can note that the performances of NSA and QNSA are very close, and in some situation it outperforms the NLMS.

4. THE VARIABLE STEP SIZE-QUANTIZED NORMALIZING FACTOR PRE-WHITENED SIGN ALGORITHM (VSS-QN-PSA)

4.1 The proposed algorithm

Since, we aim the application of acoustic echo cancellation, the step size is chosen to be varying according to the amount of the echo. The proposed algorithm is the result of the ideas presented in the previous sections, and it is resumed as follows:

$$\begin{cases} H(k+1) = H(k) + \dfrac{\mu(k) \operatorname{Sign}(e(k))}{Q\left(\sum_{i=0}^{L-1} |x^f(k-i)| + \beta_h\right)} X^f(k) \\ x^f(k) = x(k) - P(k)^T \tilde{X}(k-1) \\ P(k+1) = P(k) + \dfrac{\mu_p \operatorname{sign}\left(x^f(k)\right)}{Q\left(\sum_{i=1}^{L_p} |x(k-i)| + \beta_p\right)} \tilde{X}(k-1) \end{cases}, \quad (8)$$

The three-state step size $\mu(k) \in \{\mu_s, \mu_m, \mu_f\}$ is chosen according to the rules resumed in table1. In this table, the time parameter $D(k)$ measures the hangover that we should apply, the set of parameters $\{\tau_i\}, i = 0, .., 5$ are considered as thresholds for comparison between the amount of the far-end speech $M_x(k)$ (which yields to echo) and the amount of the error signal $M_e(k)$ composed of additive noise, residual echo and near-end speech (if it exists). $M_e(k)$ and $M_x(k)$ are evaluated in a recursive manner :

$$\begin{aligned} M_e(k) &= \gamma_e M_e(k-1) + (1-\gamma_e)\,|e(k)| \\ M_x(k) &= \gamma_x M_x(k-1)+) + (1-\gamma_x)\,|x(k)|, \end{aligned} \quad (9)$$

where γ_x and γ_e are forgetting factors chosen equal to $\gamma_x = \gamma_e = 0.996$.

The proposed decision rule depends on three parameters: the previous step size, time parameter $D(k)$, and the error magnitude compared to the magnitude of the far-end signal. In these rules we prohibit the transition from slow adaptation (double talk phase) to fast adaptation (large echo), in order to avoid abrupt adaptation change. Furthermore, when switching from slow adaptation state to medium adaptation state, we allow a hangover time to switch again from medium adaptation to fast adaptation.

Table 1: Step-sizes soft-switching rules

$\mu(k-1)$	conditions	$\mu(k)$	$D(k)$
μ_m	$\tau_1 M_x(k) < M_e(k) < \tau_2 M_x(k)$ and $D(k-1) = 0$ ms	μ_f	$D(k-1)$
μ_m	$M_e(k) > \tau_4 M_x(k)$ and $D(k-1) = 0$ ms	μ_s	$D(k-1)$
μ_m	$D > 0$ ms	μ_m	$D(k-1)$ -0.0625 ms
μ_f	$M_e(k) < \tau_0 M_x(k)$	μ_m	0 ms
μ_f	$M_e(k) > \tau_5 M_x(k)$	μ_s	$D(k-1)$
μ_s	$M_e(k) < \tau_3 M_x(k)$	μ_m	20 ms

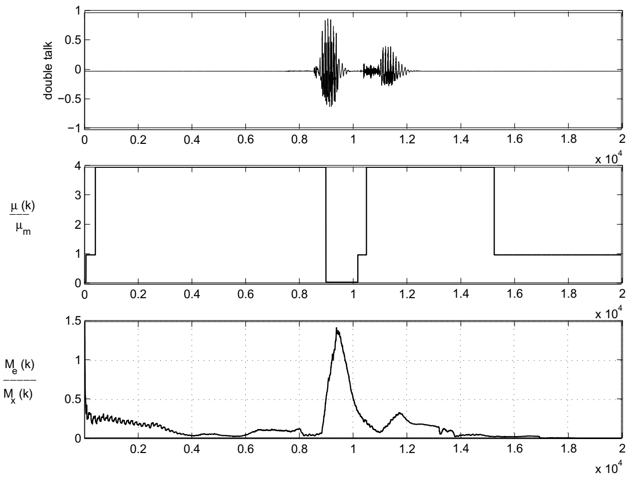

Figure 4: Variation of the step size

4.2 Simulation results

We carry out the following experiments : the system impulse response is an acoustic response of a room truncated to $L = 256$ (for a sampling frequency of 16 Khz). The microphone signal is composed of periods of only-echo and double talk. We consider the case of one-tap predictor. The step sizes are adjusted in order to insure a good trade-off between fast convergence rate and robustness during double talk.

In figure 4, we plot the double talk signal (curve 1), the variation of step sizes (curve 2), expressed in term of $\dfrac{\mu(k)}{\mu_m}$ and the variation of error magnitude compared to that of far-end signal (curve 3), expressed in term of $\dfrac{M_e(k)}{M_x(k)}$. This figure shows that during algorithm initialization, the step size is moderate, it becomes high during only-echo intervals. When double talk occurs, the step size switches to the low value. When $\dfrac{M_e(k)}{M_x(k)}$ decreases, even it is double talk (low near-end and low echo importance), the step sizes becomes moderate. We can conclude that our algorithm tracks near-end and far-end importance and adjust the step size according to that fact.

In figure 5, we report the performance of the adaptive filter through the evolution of the MSD. We compared the performance of the proposed algorithm to the NSA and the

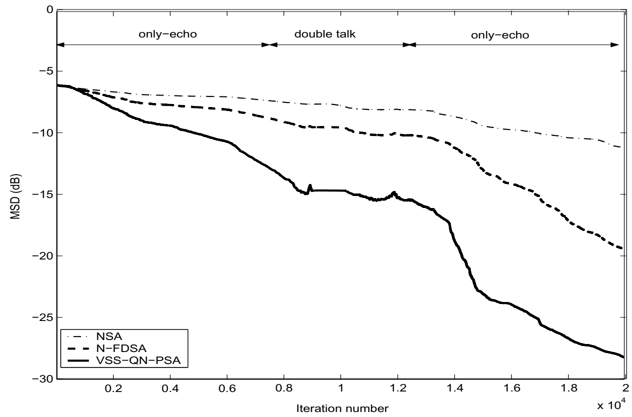

Figure 5: Performance of the proposed algorithm

normalized filtered dual sign algorithm (N-FDSA). This figure shows that the proposed (VSS-QN-PSA) algorithm accelerates the convergence rate while remaining stable during double talk.

5. CONCLUSION

With the use of the normalized sign algorithm, the introduction of pre-whitening input to it, the association of a simple predictor algorithm, the quantization of the normalization factor, and the definition of a new way of varying the step size, we have proposed a fast and robust low implementation complexity adaptive algorithm, which out-performs the classical ones in the acoustic echo cancellation field.

REFERENCES

[1] I. Gilloire, :'State of the art in Acoustic Echo Cancellation' (A. R. Figueiras and D. Docampo, 1991).

[2] S. Ben Jebara, and H. Besbes, : "On pre-whitened sign algorithms," *Proc. of the International Symposium on Circuits and Systems ISCAS*, Bangkok, Thailand, May 2003.

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

[4] C. P. Kwong, "Dual sign algorithm for adaptive filtering", *IEEE Trans. Comm.*, vol. 34, pp. 1272-1275, Dec. 1986.

[5] S. Ben Jebara, H. Besbes, "A variable step size filtered sign algorithm for acoustic echo cancellation," *IEE Electronic Letters*, vol. 39, no. 12, pp. 936-938, June 2003.

[6] S. Douglas, "A family of normalized LMS algorithms", *IEEE signal processing letters*, vol. 1, no. 3, pp. 49-51, March 1994.

[7] J. Nagumo and A. Noda, "A learning method for system identification," *IEEE Trans. Automat. Contr.,*, vol. AC-12, no. 3, pp. 282-287, June 1967.

[8] E. Eweda, "Analysis and design of a signed regressor LMS algorithm for stationary and non stationary adaptive filtering with correlated Gaussian data," *IEEE Trans. Circuits Syst.*, vol. 37, no. 11, pp. 1367-1374, Nov. 1990.