A Novel LMS Algorithm Applied to Adaptive Noise Cancellation

J. M. Górriz, Javier Ramírez, S. Cruces-Alvarez, Carlos G. Puntonet, Elmar W. Lang, and Deniz Erdogmus, *Senior Member, IEEE*

***Abstract*—In this letter, we propose a novel least-mean-square (LMS) algorithm for filtering speech sounds in the adaptive noise cancellation (ANC) problem. It is based on the minimization of the squared Euclidean norm of the difference weight vector under a stability constraint defined over the *a posteriori* estimation error. To this purpose, the Lagrangian methodology has been used in order to propose a nonlinear adaptation rule defined in terms of the product of differential inputs and errors which means a generalization of the normalized (N)LMS algorithm. The proposed method yields better tracking ability in this context as shown in the experiments which are carried out on the AURORA 2 and 3 speech databases. They provide an extensive performance evaluation along with an exhaustive comparison to standard LMS algorithms with almost the same computational load, including the NLMS and other recently reported LMS algorithms such as the modified (M)-NLMS, the error nonlinearity (EN)-LMS, or the normalized data nonlinearity (NDN)-LMS adaptation.**

***Index Terms*—Adaptive noise canceler., least-mean-square (LMS) algorithm, speech enhancement, stability constraint.**

I. INTRODUCTION

THE widely used least-mean-square (LMS) algorithm has been successfully applied to many filtering applications, including signal modeling, equalization, control, echo cancellation, biomedicine, or beamforming [1]–[3]. The typical noise cancellation scheme is shown in Fig. 1. Two distant microphones are needed for such application to capture the nature of the noise and the speech sound simultaneously. The correlation between the additive noise that corrupts the clean speech (primary signal) and the random noise in the reference input (adaptive filter input) is necessary to adaptively *cancel* the noise of the primary signal. The adjustable weights are typically determined by the LMS algorithm [3] because of its simplicity, ease of implementation, and low computational complexity. The weight update equation for the adaptive noise canceler (ANC) is

$$\boldsymbol{w}(n+1) = \boldsymbol{w}(n) + \mu e^*(n)\boldsymbol{x}(n) \tag{1}$$

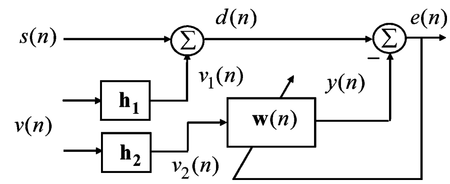

Fig. 1. Adaptive noise canceler.

where μ is a step-size parameter, $e^*(n)$ denotes the complex conjugate of the error signal $e(n)$, and $\boldsymbol{x}(n) = (x(n), \ldots, x(n-L+1))^T$ is the data vector containing L samples of the reference signal $\nu_2(n)$.

Many ANCs [3]–[6] have been proposed in the past years using modified LMS algorithms in order to simultaneously improve the tracking ability and speed of convergence. Bershad has studied the performance of the normalized LMS (NLMS) algorithm with an adaptive step size in [7] showing advantages in convergence time and steady state. Later, Douglas and Meng [6] have proposed the optimum nonlinearity for any input probability density of the independent input data samples, obtaining the normalized data nonlinearity adaptation (NDN-LMS). Although the latter algorithm is designed to improve the steady-state performance, its derivation did not consider the ANC in case of a strong target signal in the primary input. Greenberg's modified-LMS (M-LMS) [4] extended the latter approach to the case of the ANC with the nonlinearity applied to *the data vector and the target signal itself*, obtaining substantial improvements in the performance of the canceler. The disadvantage of this method is that it requires *a priori* information about the processes which is generally unknown. Recently, an interesting approach has been proposed based on a nonlinearity applied exclusively to the data vector [5].

This letter shows a novel adaptation for filtering speech signals in discontinuous speech transmission (DTX) systems, which are characterized by sudden changes of the signal statistics. The method is derived assuming stability in the sequence of *a posteriori* errors instead of the more restrictive hypothesis used in previous approaches [8], i.e., enforcing it to vanish.

II. CS-LMS ALGORITHM

The NLMS algorithm may be viewed as the solution to a constrained optimization problem [11]. The problem of interest may be stated as follows: given the tap-input vector $\boldsymbol{x}(n)$ and the desired response $d(n)$, determine the tap weight vector $\boldsymbol{w}(n+1)$ so as to minimize the squared Euclidean norm of the change $\delta\boldsymbol{w}(n+1) = \boldsymbol{w}(n+1) - \boldsymbol{w}(n)$ in the tap-weight vector $\boldsymbol{w}(n+1)$ with respect to its old value $\boldsymbol{w}(n)$, subject to the constraint $\boldsymbol{w}(n+1)^H\boldsymbol{x}(n) = d(n)$, where H denotes the Hermitian transpose. This constraint means that the *a posteriori* error sequence vanishes [$e^{[k+1]}(n) \equiv d(n) - \boldsymbol{w}(k+1)^H\boldsymbol{x}(n) = 0$, for $k = n$].

Manuscript received August 04, 2008; revised October 05, 2008. Current version published December 12, 2008. This work was supported in part by the PETRI DENCLASES (PET2006-0253), TEC2007-68030-C02-01, and TEC2008-02113 projects of the Spanish MEC and in part by the Excellence Project (TIC-02566) of the Consejería de Innovación, Ciencia y Empresa (Junta de Andalucía, Spain). This work was written in English with the help of Mrs. M. Eugenia Cobo Lara. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Fernando Perez-Gonzalez..

The authors are with the Department of Signal Theory, Networking, and Communications, University of Granada, Granada 18071, Spain (e-mail: gorriz@ugr.es).

Digital Object Identifier 10.1109/LSP.2008.2008584

In order to solve this optimization problem, the method of Lagrange multipliers is used with the Lagrangian function

$$\mathcal{L}\left(\boldsymbol{w}(n+1)\right) = \|\delta\boldsymbol{w}(n+1)\|^2 + Re\left[\lambda^* e^{[n+1]}(n)\right] \qquad (2)$$

where λ^* is the Lagrange multiplier, thus obtaining the well-known adaptation rule in (1) with the normalized step size given by $\mu = \hat{\mu}/\|\boldsymbol{x}(n)\|^2$. The latter constraint is overly restrictive in real applications; thus, if we relax it, another interesting solution can be derived. Consider the constrained optimization problem that provides the following cost function:

$$\mathcal{L}\left(\boldsymbol{w}(n+1)\right) = \|\delta\boldsymbol{w}(n+1)\|^2 + Re\left[\lambda^* \delta e^{[n+1]}(n)\right] \qquad (3)$$

where $\delta e^{[n+1]}(n) \equiv e^{[n+1]}(n) - e^{[n+1]}(n-1)$. This equilibrium constraint ensures *stability* in the sequence of *a posteriori errors*, i.e., the optimal solution $\boldsymbol{w}^{opt}(n+1)$ is the one that renders the sequence of errors as smooth as possible. Taking the partial derivative of (3) with respect to the vector $\boldsymbol{w}^H(n+1)$ and setting it equal to zero leads to

$$\begin{aligned}\frac{\partial\mathcal{L}\left(\boldsymbol{w}(n+1)\right)}{\partial\boldsymbol{w}^H(n+1)} &= \frac{\partial\delta\boldsymbol{w}^H(n+1)\delta\boldsymbol{w}(n+1)}{\partial\boldsymbol{w}^H(n+1)} + \frac{\partial}{\partial\boldsymbol{w}^H(n+1)}\\ &\quad\times Re\left[\lambda^*\left(e^{[n+1]}(n) - e^{[n+1]}(n-1)\right)\right]\\ &= 0. \end{aligned} \qquad (4)$$

Since $e^{[n+1]}(k) = d(k) - \boldsymbol{w}^H(n+1)\boldsymbol{x}(k)$ for $k = n, n-1$ and $Re[z] = 1/2(z + z^*)$, then

$$\frac{\partial\mathcal{L}\left(\boldsymbol{w}(n+1)\right)}{\partial\boldsymbol{w}^H(n+1)} = \delta\boldsymbol{w}(n+1) - \frac{1}{2}\lambda^*\delta\boldsymbol{x}(n) = 0 \qquad (5)$$

where $\delta\boldsymbol{x}(n) = \boldsymbol{x}(n) - \boldsymbol{x}(n-1)$ is the difference between two consecutive input vectors. Hence, the step of the algorithm is

$$\delta\boldsymbol{w}(n+1) = \frac{1}{2}\lambda^*\delta\boldsymbol{x}(n) \Rightarrow \boldsymbol{w}(n+1) = \boldsymbol{w}(n) + \frac{1}{2}\lambda^*\delta\boldsymbol{x}(n). \qquad (6)$$

Finally, after multiplying both sides of (5) by $\delta\boldsymbol{x}^H(n)$, the Lagrange multiplier can be expressed as

$$\lambda^* = \frac{2\delta\boldsymbol{x}^H(n)\delta\boldsymbol{w}(n+1)}{\|\delta\boldsymbol{x}(n)\|^2} = -\frac{2\left(\delta e^{[n+1]}(n) - \delta e^{[n]}(n)\right)^*}{\|\delta\boldsymbol{x}(n)\|^2} \qquad (7)$$

where $\delta e^{[n]}(n) = e^{[n]}(n) - e^{[n]}(n-1)$ is the difference in the *a priori* error sequence [denoted by $\delta e(n)$ for short], since the numerator on the left-hand side of (7) is equal to $\boldsymbol{x}^H(n)\boldsymbol{w}(n+1) - \boldsymbol{x}^H(n-1)\boldsymbol{w}(n+1) - \boldsymbol{x}^H(n)\boldsymbol{w}(n) + \boldsymbol{x}^H(n-1)\boldsymbol{w}(n)$. Therefore, applying the equilibrium constraint on the right-hand side of (7) ($\delta e^{[n+1]}(n) = 0$) leads to

$$\lambda = \frac{2\delta e^{[n]}(n)}{\|\delta\boldsymbol{x}(n)\|\,2}. \qquad (8)$$

Finally, the minimum of the Lagrangian function satisfies the following constrained stability update condition (CS-LMS)

$$\boldsymbol{w}(n+1) = \boldsymbol{w}(n) + \frac{\delta\boldsymbol{x}(n)\delta e^*(n)}{\|\delta\boldsymbol{x}(n)\|^2}. \qquad (9)$$

The weight adaptation rule can be made more robust by introducing a small positive constant ϵ into the denominator to prevent numerical instabilities in case of a vanishingly small squared norm $\|\delta\boldsymbol{x}(n)\|^2$ and by multiplying the weight increment by a constant step size μ to control the speed of the adaptation. Note that the equilibrium condition enforces the convergence of the algorithm if $\|\delta\boldsymbol{x}(n)\|^2 \neq 0$. Several learning algorithms, where the learning relies on the concurrent change of processing variables, have been proposed in the past for decorrelation, blind source separation, or deconvolution applications [9]. Stochastic information gradient (SIG) algorithms [9] maximize (or minimize) the Shannon's entropy of the sequence of errors using an estimator based on an instantaneous value of the probability density function (pdf) and Parzen windowing. In this way, the CS-LMS algorithm can be considered as a generalization of the single sample-based SIG algorithm using variable kernel density estimators [10].

III. Theoretical Remarks on the CS-LMS Adaptation

Once the CS-LMS method has been derived, a comparison is established with the NLMS algorithm. This section shows that, under some conditions: 1) CS-LMS and NLMS algorithms converge to the optimal Wiener solution $\boldsymbol{w}_o$, and 2) for any fixed step size μ, the proposed CS-LMS exhibits improvements in excess minimum squared error (EMSE) and misadjustment (M) [11] when compared to the NLMS algorithm.

A. Convergence Analysis of CS-LMS

Theorem 1 (Convergence Equivalence): Let $\boldsymbol{x}(n)$ be the tap inputs to a transversal filter and $\boldsymbol{w}(n)$ the corresponding tap weights. The estimation error $e(n)$ is obtained by comparing the estimate $y(n)$ provided by the filter with the desired response $d(n)$, that is, $e(n) = d(n) - y(n)$. On the other hand, if the desired signal $d(n)$ is generated by the multiple linear regression model, i.e., $d(n) = \boldsymbol{w}_o^H\boldsymbol{x}(n) + e_o(n)$, where $e_o(n)$ is an uncorrelated white-noise process that is statistically independent of the input vector $\boldsymbol{x}(n)$, then the CS-LMS adaptation converges to the Wiener solution $\boldsymbol{w}_o$ under stationary environment.

Proof: This theorem is proven by showing that $\boldsymbol{w}'_o = \arg\min_{\boldsymbol{w}} E[|\delta e(n)|^2]$ is equal to $\arg\min_{\boldsymbol{w}} E[|e(n)|^2] = \boldsymbol{w}_o$. This condition is satisfied since the cross-correlation vector between the concurrent change in the desired responses (δd) and input-vectors $(\delta\boldsymbol{x})$, $\boldsymbol{r}_{\delta d\delta\boldsymbol{x}} \equiv E[\delta\boldsymbol{x}\delta d^*] = \boldsymbol{R}_{\delta x}\boldsymbol{w}_o$, where $\boldsymbol{R}_{\delta x} \equiv E[\delta\boldsymbol{x}\delta\boldsymbol{x}^H]$ denotes auto-correlation matrix of $\delta\boldsymbol{x}$.

B. Learning Curves of the CS-LMS Algorithm: EMSE and Misadjustment

It is common in practice to use ensemble-average learning curves to study the statistical performance of adaptive filters. The derivation of these curves is slightly different for the ANC problem due to the presence of the desired clean signal $s(n)$. Using the definition of the weight-error vector $\boldsymbol{\varepsilon}(n) = \mathbf{w}_o - \mathbf{w}(n)$ and (9) with the step size defined as μ, we may express the evolution of $\boldsymbol{\varepsilon}(n)$ as

$$\begin{aligned}\boldsymbol{\varepsilon}(n+1) &= \boldsymbol{\varepsilon}(n) - \mu\delta\boldsymbol{x}(n)\\ &\quad\times\left(\delta s(n) + \delta v(n) - \left(\mathbf{w}_o - \boldsymbol{\varepsilon}(n)\right)^H\delta\boldsymbol{x}(n)\right)^* \end{aligned} \qquad (10)$$

where $\delta[.](n) = [.](n) - [.](n-1)$ and $v(n)$ denotes the noise in the primary signal $d(n)$ (ν_1 in Fig. 1). If $v(n)$ is assumed to be generated by the multiple regression model: $v(n) = \boldsymbol{w}_o^H\boldsymbol{x}(n) + e_o(n)$, the weight-error vector is expressed as

$$\boldsymbol{\varepsilon}(n+1) = \left(\mathbf{I} - \mu\delta\boldsymbol{x}(n)\delta\boldsymbol{x}(n)^H\right)\boldsymbol{\varepsilon}(n)$$

$$-\mu\delta\boldsymbol{x}(n)\left(\delta e_o(n)+\delta s(n)\right)^*. \quad (11)$$

By invoking the direct-averaging method [11], the equation above leads to

$$\boldsymbol{\varepsilon}_o(n+1)=(I-\mu\mathbf{R}_{\delta\boldsymbol{x}})\boldsymbol{\varepsilon}_o(n)-\mu\delta\boldsymbol{x}(n)\delta\tilde{e}_o^*(n) \quad (12)$$

where $\delta\tilde{e}_o(n)=\delta e_o(n)+\delta s(n)$, and the mean-squared error produced by the filter is given by

$$J(n)=J_o+E\left[|s(n)|^2\right]+E\left[\boldsymbol{\varepsilon}_o^H(n)\boldsymbol{x}(n)\boldsymbol{x}(n)^H\boldsymbol{\varepsilon}_o(n)\right] \quad (13)$$

where $J_o=E[|e_o(n)|^2]$ and $J_{min}=J_o+E[|s(n)|^2]$.

The stochastic evolution on the natural modes can be studied by transforming (12) into

$$\mathbf{v}(n+1)=(I-\mu\mathbf{\Lambda})\mathbf{v}(n)-\boldsymbol{\phi}(n) \quad (14)$$

and by applying the unitary similarity transformation [11] to the correlation matrix $\mathbf{R}_{\delta\boldsymbol{x}}$, where $\mathbf{\Lambda}=\mathbf{Q}^H\mathbf{R}_{\delta\boldsymbol{x}}\mathbf{Q}$ is a diagonal matrix consisting of the eigenvalues λ_k of $\mathbf{R}_{\delta\boldsymbol{x}}$, $\mathbf{Q}$ is a unitary matrix whose columns constitute an orthogonal set of eigenvectors and the stochastic force vector is defined as $\boldsymbol{\phi}(n)=\mu\mathbf{Q}^H\delta\boldsymbol{x}(n)\delta\tilde{e}_o^*(n)$. This vector has the following properties.

- The mean of the stochastic force vector $\phi(n)$ is zero: $E[\boldsymbol{\phi}(n)]=\mathbf{0}$.
- The correlation matrix of the stochastic force vector is a diagonal matrix: $E[\boldsymbol{\phi}(n)\boldsymbol{\phi}^H(n)]=\mu^2\tilde{J}\mathbf{\Lambda}$, where $\tilde{J}=2(E[|e_o(n)^2|]+E[|s(n)^2|]-Re\{r_s(1)\})$, and $r_s(1)\equiv E[s^*(n+1)s(n)]$.

The first two moments of the natural modes $\mathbf{v}(n)$ can be obtained by using these properties as in [11], which allow one to show the evolution of $J(n)$ with time step n. The third term of (13), in light of the direct-averaging method, is equal to

$$\begin{aligned}J_{ex}(n)&=E\left[\boldsymbol{\varepsilon}_o^H(n)\boldsymbol{x}(n)\boldsymbol{x}(n)^H\boldsymbol{\varepsilon}_o(n)\right]\\&\simeq E\left[\boldsymbol{\varepsilon}_o^H(n)\mathbf{R}\boldsymbol{\varepsilon}_o(n)\right]=tr\left\{\mathbf{R}E\left[\boldsymbol{\varepsilon}_o(n)\boldsymbol{\varepsilon}_o^H(n)\right]\right\}\\&=E\left[tr\{\mathbf{RQvv}^H\mathbf{Q}^H\}\right]=E\left[tr\{\mathbf{v}^H\mathbf{Q}^H\mathbf{RQv}\}\right]\\&=E\left[tr\left\{\mathbf{v}^H\mathbf{Q}^H\left(\frac{1}{2}\mathbf{R}_{\delta\boldsymbol{x}}+Re\left\{\mathbf{R}(1)\right\}\right)\mathbf{Qv}\right\}\right]\\&=\frac{1}{2}\sum_{k=1}^{L}\lambda_kE\left[|v_k(n)|^2\right]\\&\quad+E\left[tr\left\{\mathbf{v}^H\mathbf{Q}^HRe\left\{\mathbf{R}(1)\right\}\mathbf{Qv}\right\}\right]\end{aligned} \quad (15)$$

where $\mathbf{R}(1)=E[\boldsymbol{x}(n+1)\boldsymbol{x}^H(n)]$. Assuming that the input signal is weakly correlated ($\mathbf{R}(1)\sim\mathbf{0}$), the second term can be bounded in the last equality of (15) with the first term (natural evolution), i.e., $E[tr\{\mathbf{v}^H\mathbf{Q}^HRe\{\mathbf{R}(1)\}\mathbf{Qv}\}]\le(1/2)\sum_{k=1}^{L}\lambda_kE[|v_k(n)|^2]$, and then

$$\begin{aligned}J_{ex}(n)&\le\sum_{k=1}^{L}\lambda_kE\left[|v_k(n)|^2\right]\\&=\sum_{k=1}^{L}\lambda_k\left(\frac{\mu\tilde{J}}{2-\mu\lambda_k}+(1-\mu\lambda_k)^{2n}\right.\\&\qquad\left.\times\left(|v_k(0)|^2-\frac{\mu\tilde{J}}{2-\mu\lambda_k}\right)\right)\end{aligned} \quad (16)$$

where $v_k(n)$ denotes the kth-component of natural mode $\mathbf{v}(n)$ [11]. If the exponential factor is neglected with increasing n

$$J_{ex}(\infty)\le\sum_{k=1}^{L}\lambda_k\left(\frac{\mu\tilde{J}}{2-\mu\lambda_k}\right)\simeq\frac{1}{2}\mu\tilde{J}tr\{\mathbf{R}_{\delta\boldsymbol{x}}\}. \quad (17)$$

The reduction in $J_{ex}(\infty)$ is achieved whenever

$$\begin{aligned}J_{ex}(\infty)&\simeq\frac{1}{2}\mu\tilde{J}tr\{\mathbf{R}_{\delta\boldsymbol{x}}\}\simeq\mu\tilde{J}tr\{\mathbf{R}\}\\&\le J_{ex}^{\mathrm{LMS}}(\infty)\simeq\frac{1}{2}\mu J_{min}tr\{\mathbf{R}\}\Leftrightarrow Re\left\{r_s(1)\right\}\\&\ge\frac{3}{4}J_{min}\end{aligned} \quad (18)$$

i.e., the desired signal is strongly correlated. It also follows from classical analysis [11] that 1) the high value of μ balances the trade-off between $J_{ex}(\infty)$ and the average time constant τ since

$$\tau\simeq\frac{L}{\mu tr\{\mathbf{R}_{\delta\boldsymbol{x}}\}} \quad (19)$$

where L is the filter length, and 2) a necessary condition for stability is that $0<\mu<2/\lambda_k$, for all k.

IV. Experiments

The experimental analysis is mainly focused on the determination of EMSE,[1] and the misadjustment at different SNR levels, step sizes, and environments, since these quantities perfectly define the filtering performance of the algorithm. The impulse responses of the filters $\mathbf{h_1}$ and $\mathbf{h_2}$ were modeled, for practical reasons, as low-pass IIR filters

$$\mathrm{H}_1^{-1}(z)=1-0.3z^{-1}-0.1z^{-2},\quad \mathrm{H}_2^{-1}(z)=1-0.2z^{-1} \quad (20)$$

and $\epsilon=0.0001$.

A. Numerical Experiment

The first evaluation experiments considered a simple ANC configuration to test the analytical results shown in Section III. In this case, the desired signal $s(n)$ is a sum of an intermittent zero-mean AR(1) process with variance 1 and its pole at $a_s(1)=0.99$ and a zero-mean additive white noise with variance 0.001. The AR(1) process turns on and off every 3000 samples. The noise source $v(n)$ is a zero-mean Gaussian process with variance 1, and it is assumed to be independent of $s(n)$. Both CS-LMS and NLMS algorithms use an eight-tap weight vector initialized to zero and different step sizes (0.001, 0.01, 0.1). The Monte Carlo simulations resulting of running the two algorithms (over 100 trials) are shown in Fig. 2 for $\mu=0.1$. It is shown that $J(\infty)$ and τ of the CS-LMS algorithm are larger than for the NLMS method over noise segments as expected: using (13) $J_{\mathrm{LMS}}(\infty)\simeq-28.4873$ dB; $J_{\mathrm{CS-LMS}}(\infty)\simeq-27.7641$ dB. Note that $J_{min}=-30$ dB on noise segments. Somehow when $s(n)$ turns on (to model correlated speech segments), there is a clear reduction in EMSE if the value of μ is sufficiently high (to cope with a nonstationary environment). However, on speech segments: $J_{\mathrm{LMS}}(\infty)\simeq1.6012$ dB; $J_{\mathrm{CS-LMS}}(\infty)\simeq0.074$ dB.

B. Nonstationary Environment

To check the tracking ability and the robustness against noise of the proposed algorithm, let us assume that the noise ν in Fig. 1

[1] $\mathrm{EMSE}(k)=(1/J)\sum_{j=1}^{J}|e(k-j)-s(k-j)|^2$, where $J=200$ is the number of samples used in the estimation

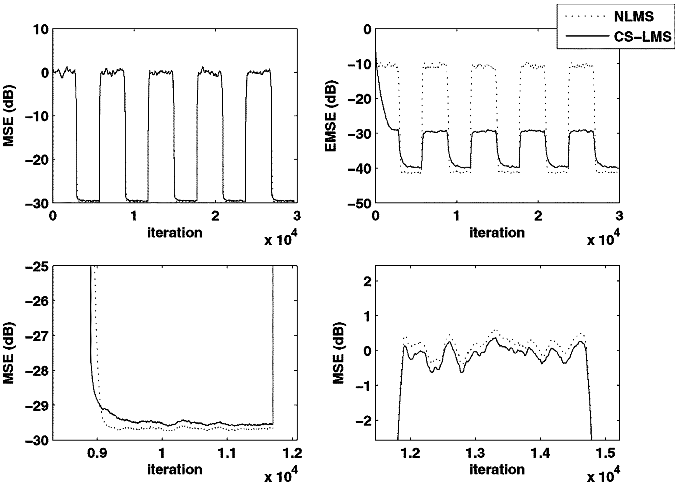

Fig. 2. Numerical experiment in the ANC problem. Top: MSE and EMSE [in $\mathrm{dB} = 10\log_{10}(.)$] comparison between the CS-LMS and NLMS algorithms. Bottom: zoom on MSE evolution over noise and speech segments. Stationary environment ($\mu = 0.1, \sigma_v^2 = 1$ $L = 8$).

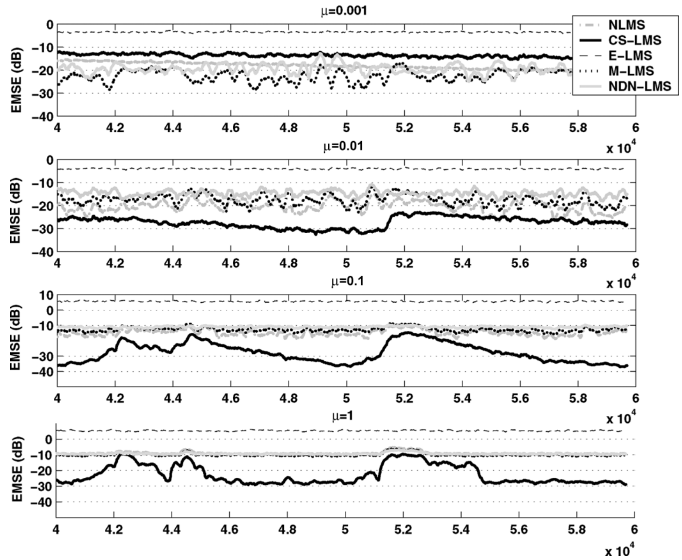

Fig. 3. EMSE plots in steady-state averaged over an utterance (50 realizations, set of step sizes μ, $L = 12$).

is one of the eight real-world noises extracted from AURORA 2 database at SNRs from 20 dB to -5 dB [13], i.e., babble noise. In this case, $s(n)$ was selected from the AURORA subset of the original Spanish SpeechDat-Car database [12], which contains 4914 clean recordings from more than 160 speakers. Several experiments are obtained by varying the filter length $L = 8;12;24$ and the step size $\mu = \{0.001; 0.01; 0.1; 1\}$ of the algorithms according to DTX application. Note that μ should be large enough to cope with rapid transitions in the channel. The range for the step size μ was selected empirically.

Fig. 3 shows the operation of the algorithms for filtering a speech signal corrupted by noise in a DTX scenario. Observe how the equilibrium constraint obtains the best trade-off in the filtering performance of the canceler. Finally, Table I summarizes the *averaged results* of the EMSE and M in a nonstationary environment using the proposed and referenced algorithms for all the recordings of the database and the set of parameters $\{\mu, L, \mathrm{SNRs}\}$ and noises. Thus, we are including the optimal results of each algorithm and check which one obtains the best averaged accuracy. The proposed method yields the minimum EMSE and M for the selected range of filter lengths and step sizes as shown in Table I.

TABLE I
PERFORMANCE OF REFERENCED AND PROPOSED LMS ALGORITHMS IN NONSTATIONARY ENVIRONMENT

Noise		N-LMS		NDN-LMS	
		$\overline{\text{EMSE}}$(dB)	$\overline{\text{M}}$	$\overline{\text{EMSE}}$(dB)	$\overline{\text{M}}$
Aurora 2:	Babble, etc.	-12.46	1.13	-8.43	2.46
CS-LMS		**EN-LMS**		**M-LMS**	
$\overline{\text{EMSE}}$(dB)	$\overline{\text{M}}$	$\overline{\text{EMSE}}$(dB)	$\overline{\text{M}}$	$\overline{\text{EMSE}}$(dB)	$\overline{\text{M}}$
-17.58	0.62	-10.13	2.32	-12.90	1.10

V. CONCLUSION

This letter introduced a novel CS-LMS algorithm based on the concept of difference quantities and the constraint of equilibrium condition in the sequence of *a posteriori estimation errors*. The method, which applies nonlinearities to the error and input signal sequences, was derived using the Lagrange multiplier method as a generalization of the NLMS algorithm. Under certain conditions, the proposed ANC based on the CS-LMS algorithm showed improved performance by decreasing the excess mean-squared error and misadjustment compared to referenced algorithms [4]–[7].

REFERENCES

[1] S. J. Elliott and P. A. Nelson, "Active noise control," *IEEE Signal Process. Mag.*, vol. 10, no. 4, pp. 12–35, Oct. 1993.
[2] V. E. DeBrunner and D. Zhou, "Hybrid filtered error LMS algorithm: Another alternative to F-XLMS," *IEEE Trans. CircuSyst.*, vol. 53, no. 3, pp. 653–661, Mar. 2006.
[3] B. Widrow, J. R. Glover, J. M. Mccool, J. Kaunitz, C. S. Williams, R. H. Hean, J. R. Zeidler, E. Dong, and R. C. Goodlin, "Adaptive noise cancelling: Principles and applications," *Proc. IEEE*, vol. 63, no. 12, pp. 1692–1716, Dec. 1975.
[4] J. E. Greenberg, "Modified LMS algorithms for speech processing with an adaptive noise canceler," *IEEE Trans. Speech Audio Process.*, vol. 6, no. 4, pp. 338–351, Jul. 1998.
[5] Z. Ramadan and A. Poularikas, "An adaptive noise canceler using error nonlinearities in the LMS adaptation," in *Proc. IEEE SoutheastCon 2004*, Mar. 2004, vol. 1, pp. 359–364.
[6] S. C. Douglas and T. H. Y. Meng, "Normalized data nonlinearities for LMS adaptation," *IEEE Trans. Signal Process.*, vol. 42, no. 6, pp. 1352–1354, Jun. 1994.
[7] N. J. Bershad, "Analysis of the normalized LMS algorithm with Gaussian inputs," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 34, no. 4, pp. 793–806, Aug. 1986.
[8] S. C. Douglas, "A family of normalized LMS algorithms," *IEEE Signal Process. Lett.*, vol. 1, no. 3, pp. 49–51, Mar. 1994.
[9] D. Erdogmus, K. E. Hild, and J. C. Principe, "Online entropy manipulation: Stochastic information gradient," *IEEE Signal Process. Lett.*, vol. 10, no. 8, pp. 242–245, Aug. 2003.
[10] L. Breiman, W. Meisel, and E. Purcell, "Variable kernel estimates of probability densities," *Technometrics*, vol. 19, pp. 135–144, 1977.
[11] S. Haykin, *Adaptive Filter Theory*. Englewood Cliffs, NJ: Prentice-Hall, 1986.
[12] A. Moreno, B. Lindberg, C. Draxler, G. Richard, K. Choukri, S. Euler, and J. Allen, "SpeechDat-Car, a large speech database for automotive environments," in *Proc. Int. Conf. Language Resources Evaluation (LREC)*, May 2000.
[13] H. Hirsch and D. Pearce, "The AURORA experimental framework for the performance evaluation of speech recognition systems under noise conditions," in *Proc. Automatic Speech Recognition: Challenges Next Millennium (ISCA)*, Sep. 2000.