

Generalized Optimal Step-Size for Blind Multichannel LMS System Identification

Nikolay D. Gaubitch, *Student Member, IEEE*, Md. Kamrul Hasan, *Senior Member, IEEE*, and Patrick A. Naylor, *Member, IEEE*

Abstract—The choice of step-size in adaptive blind channel identification using the multichannel least mean squares (MCLMS) algorithm is critical and controls its convergence rate, stability, and sensitivity to noise. In this letter, we derive the expression for an optimal step-size in the Wiener sense and investigate its properties. An implementation technique for the Wiener solution of the self-adaptive step-size is presented, and it is shown that significant performance improvements are obtained compared to existing approaches in the presence of noise.

Index Terms—Blind channel identification (BCI), multichannel least mean square (MCLMS) algorithm, Wiener step-size.

I. INTRODUCTION

BLIND channel identification (BCI) has several applications in various fields of engineering, in particular where blind deconvolution or source separation is required. Examples include communications systems where the received signal must be equalized to obtain the transmitted signal [1], geophysics where the reflectivity of the earth layers is explored by extracting seismic signals from the sensor observations [2], and speech dereverberation where the acoustic impulse responses are estimated blindly from reverberant speech, and then deconvolution is performed to remove the effects of the room [3].

Several blind multichannel identification algorithms have been reported in the literature. A review of many existing methods can be found in [4]. Recently, a class of adaptive approaches based on the cross-relation between channels [5] was proposed, with implementations in the time-domain [6] and in the frequency-domain [7]. Such algorithms are attractive for real-time applications with the simplest method being multichannel least mean squares (MCLMS). One of the drawbacks of MCLMS is the difficulty in choosing the step-size for adaptation that governs the rate of convergence and the steady-state behavior. Furthermore, selection of the step-size normally depends on the input signal power. Therefore, it is of interest to find a self-adaptive step-size that is optimal in some sense.

Previous work on nonblind adaptive system identification has shown that deployment of variable and optimal step-sizes results

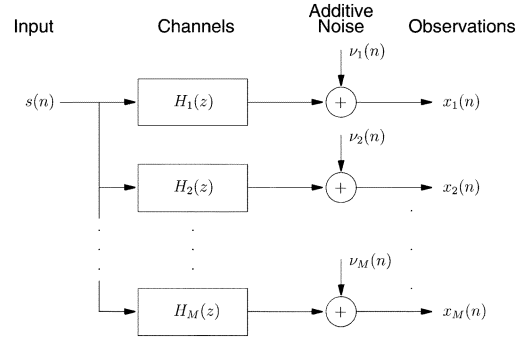


Fig. 1. SIMO system block diagram.

in performance exceeding that of a fixed step-size (see, e.g., [8] and [9]). The first approach for optimal variable step-size in BCI was presented in [10] for the unconstrained MCLMS, where the optimal step-size is obtained at each iteration using the instantaneous values of the channel and gradient estimates.

In this letter, we derive a general optimal step-size for the unconstrained MCLMS algorithm, extending the previous approach to both noisy and noise-free cases. In a noise-free environment, the optimal step-size in [10] becomes a special case of our proposed Wiener step-size. In the presence of noise, an underlying assumption in [10] is invalid, and we show that, in such cases, our optimal step-size can give a significant improvement.

The remainder of the letter is organized as follows. In Section II, the problem of blind channel estimation is formulated, and a summary of the MCLMS algorithm is given in Section III. Section IV derives the optimal step-size, and a technique for its implementation is discussed. Simulation results are provided in Section V, and finally, conclusions are drawn in Section VI.

II. BLIND CHANNEL IDENTIFICATION PROBLEM

In the single-input multiple-output (SIMO) system of Fig. 1, a signal, $s(n)$, is observed in a noisy multipath environment by an array of sensors at a distance from the source. The signal received at the i th sensor is

$$x_i(n) = \mathbf{h}_i^T \mathbf{s}(n) + \nu_i(n) \quad (1)$$

where $\mathbf{h}_i = [h_{i,0} \ h_{i,1} \ \dots \ h_{i,L-1}]^T$ is the L -tap impulse response of the channel between the source and the i th sensor, $\mathbf{s}(n) = [s(n) \ s(n-1) \ \dots \ s(n-L+1)]^T$ is the source signal vector, and $\nu_i(n)$ is measurement noise at the i th sensor.

The aim of a blind channel identification algorithm is to form an estimate $\hat{\mathbf{h}}_i = [\hat{h}_{i,0} \ \hat{h}_{i,1} \ \dots \ \hat{h}_{i,L-1}]^T$ of the impulse responses \mathbf{h}_i , using only the observations $x_i(n)$, $i = 1, 2, \dots, M$. This has been shown possible provided that the following identifiability conditions are satisfied [5]: 1) the channels do not

Manuscript received December 21, 2005; revised March 24, 2006. This work was supported by the Engineering and Physical Sciences Research Council, U.K. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Israel Cohen.

P. A. Naylor and N. D. Gaubitch are with the Department of Electrical and Electronic Engineering, Imperial College London, London SW7 2AZ, U.K. (e-mail: ndg@imperial.ac.uk; p.naylor@imperial.ac.uk).

M. K. Hasan is with the Department of Electrical and Electronic Engineering, Bangladesh University of Engineering and Technology, Dhaka-1000, Bangladesh (e-mail: khasan@eee.buet.ac.bd).

Digital Object Identifier 10.1109/LSP.2006.876341

share any common zeros, and 2) the autocorrelation matrix of the source signal is of full rank.

III. MCLMS ALGORITHM FOR BCI

The multichannel LMS algorithm proposed in [6] and [10] is based on the cross-relation between two channels, which in the noise-free case is [5]

$$\mathbf{x}_i^T(n)\mathbf{h}_j = \mathbf{x}_j^T(n)\mathbf{h}_i, \quad i, j = 1, 2, \dots, M \quad (2)$$

where $\mathbf{x}_i(n) = [x_i(n) \ x_i(n-1) \ \dots \ x_i(n-L+1)]$ is the i th observation vector. Taking all sensor pairs into account, a system of equations can be written

$$\mathbf{R}\mathbf{h} = \mathbf{0} \quad (3)$$

where $\mathbf{h} = [\mathbf{h}_1^T \ \mathbf{h}_2^T \ \dots \ \mathbf{h}_M^T]^T$ is a vector of the concatenated impulse responses, and

$$\mathbf{R} = \begin{bmatrix} \sum_{i \neq 1} \mathbf{R}_{x_i x_i} & -\mathbf{R}_{x_2 x_1} & \dots & -\mathbf{R}_{x_M x_1} \\ -\mathbf{R}_{x_1 x_2} & \sum_{i \neq 2} \mathbf{R}_{x_i x_i} & \dots & -\mathbf{R}_{x_M x_2} \\ \vdots & \vdots & \ddots & \vdots \\ -\mathbf{R}_{x_1 x_M} & -\mathbf{R}_{x_2 x_M} & \dots & \sum_{i \neq M} \mathbf{R}_{x_i x_i} \end{bmatrix} \quad (4)$$

is a correlation-like matrix with

$$\mathbf{R}_{x_i x_j} = E\{\mathbf{x}_i(n)\mathbf{x}_j^T(n)\}, \quad i, j = 1, 2, \dots, M. \quad (5)$$

Thus, the impulse responses can be identified uniquely up to a scaling factor by finding the eigenvector corresponding to the smallest eigenvalue of \mathbf{R} [5].

In the presence of noise, the cross-relation in (2) does not hold, and an error function can be defined as [6]

$$e_{ij}(n) = \mathbf{x}_i^T(n)\hat{\mathbf{h}}_j - \mathbf{x}_j^T(n)\hat{\mathbf{h}}_i \quad (6)$$

and a cost function formulated as

$$J(n) = \sum_{i=1}^{M-1} \sum_{j=i+1}^M e_{ij}^2(n). \quad (7)$$

The optimal estimate of the channels is found by minimizing $J(n)$ with respect to $\hat{\mathbf{h}}$

$$\hat{\mathbf{h}}_{\text{opt}} = \arg \min_{\hat{\mathbf{h}}} E\{J(n)\}, \text{ subject to } \|\hat{\mathbf{h}}\| = 1 \quad (8)$$

where $E\{\cdot\}$ is the expectation operator. The unit norm constraint shown here is often introduced to avoid the trivial estimate $\hat{\mathbf{h}} = \mathbf{0}$. However, it was shown in [10] that the trivial solution can be avoided if the estimation vectors are initialized appropriately, and therefore, we will use the unconstrained MCLMS for the remainder of this letter.

The LMS adaptive algorithm finds the desired solution using [11]

$$\hat{\mathbf{h}}(n+1) = \hat{\mathbf{h}}(n) - \mu \nabla J(n) \quad (9)$$

where ∇ is the gradient operator, and μ is a positive step-size. The instantaneous gradient estimate at time n has been shown to be [6], [10]

$$\nabla J(n) = 2\tilde{\mathbf{R}}(n)\hat{\mathbf{h}}(n) \quad (10)$$

where $\hat{\mathbf{h}}(n) = [\hat{\mathbf{h}}_1^T(n) \ \hat{\mathbf{h}}_2^T(n) \ \dots \ \hat{\mathbf{h}}_M^T(n)]^T$ is the channel vector estimate at time n , and $\tilde{\mathbf{R}}(n)$ is the instantaneous estimate of the matrix \mathbf{R} at time n with $\tilde{\mathbf{R}}_{x_i x_j}(n) = \mathbf{x}_i(n)\mathbf{x}_j^T(n)$, $i, j = 1, 2, \dots, M$.

Finally, substituting (10) into (9), the update equation for the unconstrained MCLMS algorithm is

$$\hat{\mathbf{h}}(n+1) = \hat{\mathbf{h}}(n) - \mu 2\tilde{\mathbf{R}}(n)\hat{\mathbf{h}}(n). \quad (11)$$

IV. WIENER SOLUTION OF THE SELF-ADAPTIVE STEP-SIZE

We would like to find the step-size, $\mu(n)$, which minimizes the misalignment at every iteration, given the current channel estimate $\hat{\mathbf{h}}(n)$. Consequently, we define the cost function

$$J_\mu(n) = E\{\|\mathbf{h} - \alpha\hat{\mathbf{h}}(n+1)\|^2 | \hat{\mathbf{h}}(n)\} \quad (12)$$

where α is the scaling constant inherent in BCI based on the cross-relation and is assumed, for now, to be known. Minimizing $J_\mu(n)$ with respect to μ gives the optimal step-size at time n , $\mu_{\text{opt}}(n) = \arg \min_{\mu} J_\mu(n)$.

Substituting (9) into (12), the cost function becomes

$$\begin{aligned} J_\mu(n) &= E\{\|\mathbf{h} - \alpha\hat{\mathbf{h}}(n) + \alpha\mu\nabla J(n)\|^2 | \hat{\mathbf{h}}(n)\} \\ &= E\{\|\mathbf{h}\|^2 - 2\alpha\mathbf{h}^T\hat{\mathbf{h}}(n) + 2\alpha\mu\mathbf{h}^T\nabla J(n) \\ &\quad - 2\alpha^2\mu\hat{\mathbf{h}}^T(n)\nabla J(n) + \alpha^2\mu^2\|\nabla J(n)\|^2 \\ &\quad + \alpha^2\|\hat{\mathbf{h}}^T(n)\|^2 | \hat{\mathbf{h}}(n)\}. \end{aligned} \quad (13)$$

The optimal step-size in the MMSE sense is obtained from

$$\frac{\partial J_\mu(n)}{\partial \mu} = 0. \quad (14)$$

Thus, using (13) and (14), we obtain the Wiener step-size

$$\mu_{\text{opt}}(n) = \frac{\hat{\mathbf{h}}^T(n) - \frac{1}{\alpha}\mathbf{h}^T}{E\{\|\nabla J(n)\|^2 | \hat{\mathbf{h}}(n)\}} E\{\nabla J(n) | \hat{\mathbf{h}}(n)\}. \quad (15)$$

As in the derivation of the LMS algorithm, the expected values may be approximated by their instantaneous estimates. Let $\nabla J(n)$ and $\|\nabla J(n)\|^2$ be the instantaneous estimates of the conditional expectations $E\{\nabla J(n) | \hat{\mathbf{h}}(n)\}$ and $E\{\|\nabla J(n)\|^2 | \hat{\mathbf{h}}(n)\}$, respectively. The optimal step-size is then obtained by

$$\tilde{\mu}_{\text{opt}}(n) = \frac{\hat{\mathbf{h}}^T(n)\nabla J(n)}{\|\nabla J(n)\|^2} - \gamma(n) \quad (16)$$

with

$$\gamma(n) = \frac{\mathbf{h}^T\nabla J(n)}{\alpha\|\nabla J(n)\|^2}. \quad (17)$$

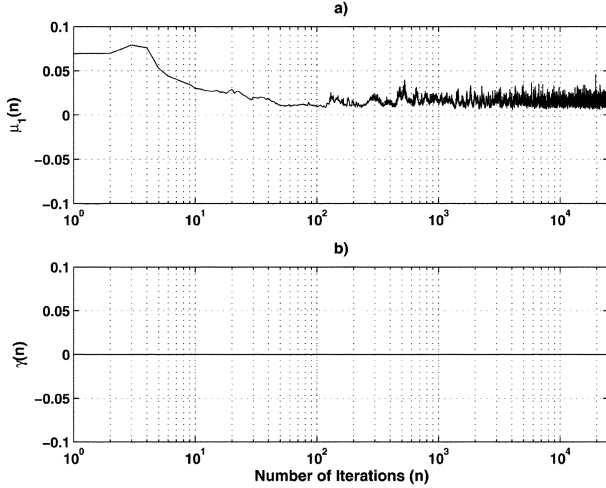


Fig. 2. Step-size behavior in the noise-free case for a) $\mu_1(n)$ and b) $\gamma(n)$, which together form $\tilde{\mu}_{\text{opt}}(n)$.

In the noise-free case, $\nu_i(n) = 0, \forall i$, and it can be assumed that the gradient vector is orthogonal to the true solution at all times, such that $\mathbf{h}^T \nabla J(n) \approx 0$, as discussed in [10]. In this case, our solution in (16) reduces to

$$\mu_1(n) = \frac{\hat{\mathbf{h}}^T(n) \nabla J(n)}{\|\nabla J(n)\|^2} \quad (18)$$

which is the result presented in [10] and can be seen as a special case of the Wiener step-size in (15). Fig. 2 shows an example plot of the optimal step-size components a) $\mu_1(n)$ from (18) and b) $\gamma(n)$ from (17), which together form $\tilde{\mu}_{\text{opt}}(n)$ (16). Note that the first 20 samples ($n = 0, 1, \dots, 19$) in Fig. 2(a) have been excluded from the plot since they are generally large and shadow the fine detail that is of interest here. It is interesting to note that the step-size $\mu_1(n)$ varies in the range 0–0.05 rather than approaching zero with convergence as might be expected. We note also that $\gamma(n)$ varies in a range very close to zero.

The assumption of orthogonality between the gradient and the true solution does not hold when noise is present, i.e., $\mathbf{h}^T \nabla J(n) \neq 0$, and the approximation employed in (18) becomes inaccurate, even when small amounts of noise are introduced. Thus, $\gamma(n)$ is not zero and becomes more significant as the SNR decreases. This can be seen in Fig. 3(b), where $\gamma(n)$ is plotted for SNR = 20 dB. The examples in Figs. 2 and 3 were generated using the simulation settings described in Section V.

In order to implement the optimal step-size in noisy conditions, we need to estimate $\gamma(n)$ as it requires knowledge of the true impulse responses that are not available. Thus, we need to form an approximate expression $\hat{\gamma}(n) \approx \gamma(n)$. It can be seen that the first term of (16), and $\gamma(n)$ will tend to equivalence as the channel estimates $\hat{\mathbf{h}}$ approach the true solution. This can also be observed in Fig. 3, where after approximately 5×10^3 iterations, $\mu_1(n)$ is close to $\gamma(n)$. Therefore, the purpose of $\gamma(n)$ is to attenuate the step-size, bringing it toward zero as the channel estimates improve. Using this as motivation, we write

$$\hat{\gamma}(n) = \beta(n) \frac{\hat{\mathbf{h}}^T(n) \nabla J(n)}{\|\nabla J(n)\|^2} \quad (19)$$

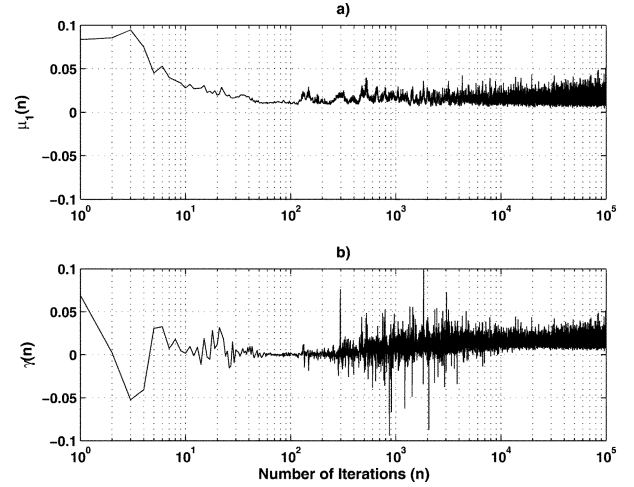


Fig. 3. Step-size behavior at SNR = 20 dB for a) $\mu_1(n)$ and b) $\gamma(n)$, which together form $\tilde{\mu}_{\text{opt}}(n)$.

where $\beta(n)$ is a weighting function that should slowly attenuate the step-size as the adaptive filter converges. Replacing $\gamma(n)$ in (16) with (19), an approximate Wiener step-size for noisy conditions is obtained

$$\mu_2(n) = (1 - \beta(n)) \frac{\hat{\mathbf{h}}^T(n) \nabla J(n)}{\|\nabla J(n)\|^2}. \quad (20)$$

A weighting function that was found to be suitable is

$$\beta(n) = \exp\left(\frac{-J(n)L\sigma_x^2}{\sigma_\nu^2}\right) \quad (21)$$

where σ_x^2 and σ_ν^2 are the observed signal power and the noise power, respectively. The expression in (21) highlights some of the desired features of the weighting function: 1) it tends to unity as the error tends to zero, 2) it tends to zero as the SNR tends to infinity, and 3) it is dependent on the error signal and therefore provides a means of tracking changes in the system.

V. SIMULATION RESULTS

Simulation results are presented to investigate the use of the step-size parameters $\tilde{\mu}_{\text{opt}}(n)$, $\mu_1(n)$, and $\mu_2(n)$ defined in Section IV. For the experiments, we used three random channels of length $L = 32$, as shown in Fig. 4. The excitation signal was white Gaussian noise, and the sampling frequency was set to $f_s = 8$ kHz. The normalized projection misalignment (NPM) [12] was used as a performance metric

$$\text{NPM}(n) = 20 \log_{10} \left(\frac{1}{\|\mathbf{h}\|} \left\| \mathbf{h} - \frac{\mathbf{h}^T \hat{\mathbf{h}}(n)}{\hat{\mathbf{h}}^T(n) \hat{\mathbf{h}}(n)} \hat{\mathbf{h}}(n) \right\| \right) \text{ dB}. \quad (22)$$

Using this measure, only the misalignment is accounted for, ignoring the effect of the arbitrary constant [12]. In all cases, the adaptive algorithms were initialized with $\hat{\mathbf{h}}(0) = [1 \ 1 \ \dots \ 1]^T$.

First, we considered the noise-free case. The unconstrained MCLMS with update equation according to (11) was executed with fixed and with variable optimal step-sizes. The results are shown in Fig. 5, where the NPM is plotted against time for a) a fixed step-size $\mu = 0.02$, b) a fixed step-size $\mu = 0.01$, c) the optimal step-size $\tilde{\mu}_{\text{opt}}(n)$, d) the optimal step-size from [10] $\mu_1(n)$, and e) the approximate optimal step-size $\mu_2(n)$ in

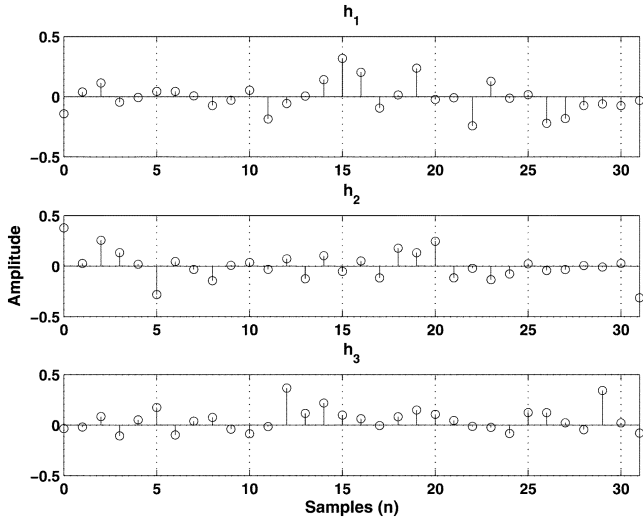


Fig. 4. Three random channels of length $L = 32$ used in the simulations.

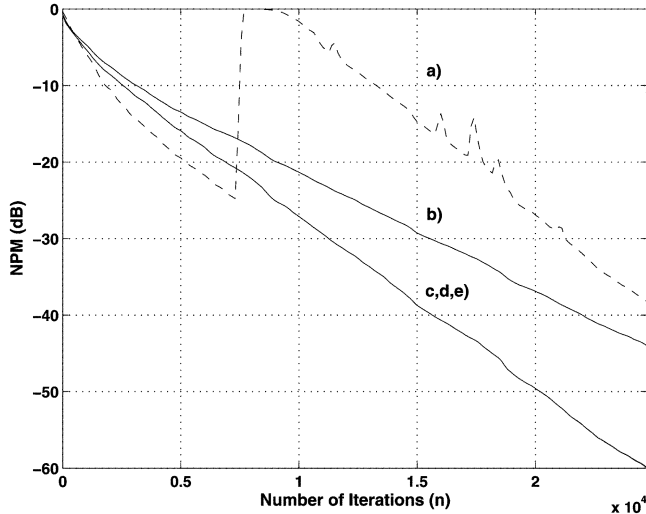


Fig. 5. Normalized projection misalignment in the noise-free case for the unconstrained MCLMS with step-size a) $\mu = 0.02$, b) $\mu = 0.01$, c) $\tilde{\mu}_{\text{opt}}(n)$, d) $\mu_1(n)$, and e) $\mu_2(n)$.

(20). Fig. 5(a) and (b) demonstrates the difficulty in choosing a fixed step-size since $\mu = 0.02$ results in instability. As expected for the noise-free case, all three implementations of the optimal step-size perform uniformly, resulting in a monotonic convergence.

Next, the case of additive noise was investigated for an example case with $\text{SNR} = 20$ dB. The three implementations of the variable step-sizes were employed with unconstrained MCLMS. The resulting NPM is plotted against time iterations n in Fig. 6 for a) the optimal step-size from [10] $\mu_1(n)$, b) the approximate optimal step-size $\mu_2(n)$ in (20), and c) the optimal step-size $\tilde{\mu}_{\text{opt}}(n)$. Two interesting observations can be made from this: 1) the implementation with the general optimal step-size $\tilde{\mu}_{\text{opt}}(n)$ converges toward zero even under noisy conditions and thus provides an indication of the obtainable NPM, and 2) using approximations like that of (20) can provide a significant improvement in terms of misalignment, as seen in Fig. 6(b).

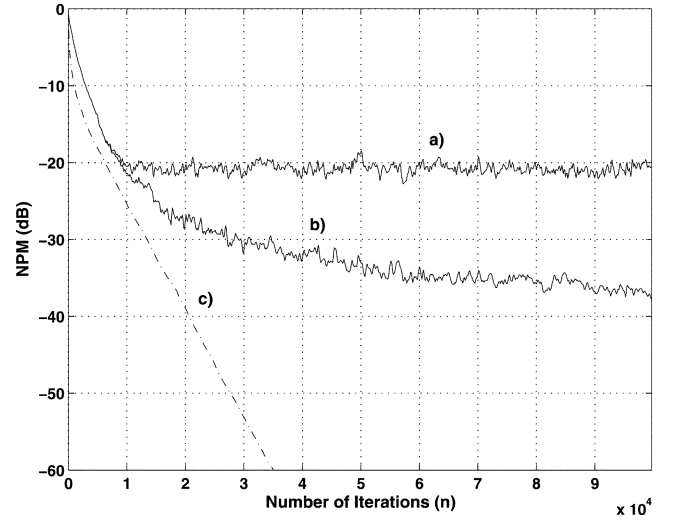


Fig. 6. Normalized projection misalignment in the noisy case ($\text{SNR} = 20$ dB) for the unconstrained MCLMS with step-size a) $\mu_1(n)$, b) $\mu_2(n)$, and c) $\tilde{\mu}_{\text{opt}}(n)$.

VI. CONCLUSION

We have derived the Wiener step-size for the unconstrained MCLMS algorithm for BCI. This step-size is calculated at every iteration. In order to implement the optimal step-size in noisy environments, we have formulated an approximate Wiener step-size. Simulation results have demonstrated that the proposed self-adaptive step-size gives significantly better NPM compared to a fixed step-size or the optimal step-size algorithm in [10].

REFERENCES

- [1] Z. Xu and M. K. Tsantsanis, "Blind channel estimation for long code multiuser CDMA systems," *IEEE Trans. Signal Process.*, vol. 48, no. 4, pp. 988–1001, Apr. 2000.
- [2] K. F. Kaarensen and T. Tøft, "Multichannel blind deconvolution of seismic signals," *Geophys.*, vol. 63, no. 6, pp. 2093–2107, Nov. 1998.
- [3] S. Gannot and M. Moonen, "Subspace methods for multi-microphone speech dereverberation," *EURASIP J. Appl. Signal Process.*, vol. 2003, no. 11, pp. 1074–1090, Oct. 2003.
- [4] L. Tong and S. Perreau, "Multichannel blind identification: From subspace to maximum likelihood methods," *Proc. IEEE*, vol. 86, no. 10, pp. 1951–1968, Oct. 1998.
- [5] G. Xu, H. Liu, L. Tong, and T. Kailath, "A least-squares approach to blind channel identification," *IEEE Trans. Signal Process.*, vol. 43, no. 12, pp. 2982–2993, Dec. 1995.
- [6] Y. Huang and J. Benesty, "Adaptive multi-channel least mean square and Newton algorithms for blind channel identification," *Signal Process.*, vol. 82, no. 8, pp. 1127–1138, Aug. 2002.
- [7] —, "A class of frequency-domain adaptive approaches to blind multichannel identification," *IEEE Trans. Signal Process.*, vol. 51, no. 1, pp. 11–24, Jan. 2003.
- [8] R. W. Harris, D. M. Chabries, and F. A. Bishop, "A variable step (VS) adaptive filter algorithm," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. ASSP-34, no. 2, pp. 309–316, Apr. 1986.
- [9] T. Aboulnasr and K. Mayyas, "A robust variable step size LMS-type algorithm: Analysis and simulations," *IEEE Trans. Signal Process.*, vol. 45, no. 3, pp. 631–639, Mar. 1997.
- [10] Y. Huang, J. Benesty, and J. Chen, "Optimal step size of the adaptive multichannel LMS algorithm for blind SIMO identification," *IEEE Signal Process. Lett.*, vol. 12, no. 3, pp. 173–176, Mar. 2005.
- [11] S. Haykin, *Adaptive Filter Theory*, 4th ed. Upper Saddle River, NJ: Prentice-Hall, 2001.
- [12] D. R. Morgan, J. Benesty, and M. M. Sondhi, "On the evaluation of estimated impulse responses," *IEEE Signal Process. Lett.*, vol. 5, no. 7, pp. 174–176, Jul. 1998.