Optimal Step Size of the Adaptive Multichannel LMS Algorithm for Blind SIMO Identification

Yiteng (Arden) Huang, *Member, IEEE*, Jacob Benesty, *Senior Member, IEEE*, and Jingdong Chen, *Member, IEEE*

***Abstract*—Adaptive algorithms for blindly identifying single-input multiple-output (SIMO) systems are appealing because of their computational efficiency and capability of continuously tracking a time-varying system. Adaptive multichannel least-mean-square (MCLMS) algorithms (with and without the unit-norm constraint) are analyzed, and the optimal step size is derived. A simple yet effective variable step-size MCLMS algorithm is proposed, and its performance is evaluated with simulations.**

***Index Terms*—Blind Channel Identification (BCI), Least Mean Square (LMS), Multichannel Signal Processing, SIMO Systems, Variable Step-Size Adaptive Algorithm.**

I. Introduction

Blind identification of single-input multiple-output (SIMO) systems has a variety of potential applications in wireless communications and other signal processing systems. As research in this area advances and demands for efficient implementation emerge, developing adaptive blind channel identification (BCI) algorithms becomes imperative and has gained increased attention in the last several years. Two important proposals, among others, in the literature are least-mean-square (LMS)-type algorithms based only on the second-order statistics of the system outputs. One is the adaptive multichannel LMS (MCLMS) algorithm (with a unit-norm constraint on the channel impulse response vector) [1], and the other is the unconstrained MCLMS (UMCLMS) algorithm [2]. Both algorithms work well for an identifiable, slowly time-varying SIMO system of moderately long channels like most wireless communication systems. However, the step size governs the rate of convergence and the steady-state misalignment error. A fixed step size usually cannot meet the conflicting requirement of fast convergence and low misalignment. Moreover, in order to prevent the algorithms from diverging, several trials need to be conducted before a proper step size is found. This drawback obviously will obstruct the use of these adaptive algorithms in practice.

In this letter, we will derive the optimal step size for the UMCLMS algorithm, which minimizes the misalignment error in each step of adaptation. Using this discovery, we will develop a variable step-size unconstrained MCLMS (VSS-UMCLMS) algorithm. The effectiveness of this step-size control scheme will be justified by simulations.

Manuscript received July 16, 2004; revised September 8, 2004. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Dominic K. C. Ho.

Y. Huang and J. Chen are with Bell Laboratories, Lucent Technologies, Murray Hill, NJ 07974, USA (e-mail: arden@research.bell-labs.com; jingdong@research.bell-labs.com).

J. Benesty is with the Université du Québec, INRS-EMT, Montréal, QC, H5A 1K6, Canada (e-mail: benesty@inrs-emt.uquebec.ca).

Digital Object Identifier 10.1109/LSP.2004.842286

II. Signal Model and Problem Formulation

Consider a SIMO finite impulse response (FIR) linear system, as depicted in Fig. 1. The ith observation $x_i(n)$ is expressed as follows:

$$x_i(n) = h_{\mathrm{t},i} * s(n) + b_i(n), \quad i = 1, 2, \cdots, M \tag{1}$$

where $s(n)$ represents the common source signal, $h_{\mathrm{t},i}$ stands for the true (subscript t) impulse response of the ith channel, $b_i(n)$ is the additive noise signal captured by the ith sensor, the symbol $*$ denotes the linear convolution operator, and M is the number of channels. In a vector/matrix form, such a relationship (1) becomes

$$\mathbf{x}_i(n) = \mathbf{H}_{\mathrm{t},i} \cdot \mathbf{s}(n) + \mathbf{b}_i(n) \tag{2}$$

where

$$\mathbf{x}_i(n) = [x_i(n)\ x_i(n-1)\ \cdots\ x_i(n-L+1)]^T$$

$$\mathbf{H}_{\mathrm{t},i} = \begin{bmatrix} h_{\mathrm{t},i,0} & h_{\mathrm{t},i,1} & \cdots & h_{\mathrm{t},i,L-1} & 0 & \cdots & 0 \\ 0 & h_{\mathrm{t},i,0} & \cdots & h_{\mathrm{t},i,L-2} & h_{\mathrm{t},i,L-1} & \cdots & 0 \\ \vdots & \ddots & \ddots & \vdots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & h_{\mathrm{t},i,0} & h_{\mathrm{t},i,1} & \cdots & h_{\mathrm{t},i,L-1} \end{bmatrix}$$

$$\mathbf{s}(n) = [s(n)\ s(n-1)\ \cdots\ s(n-L+1)\ \cdots\ s(n-2L+2)]^T$$

$$\mathbf{b}_i(n) = [b_i(n)\ b_i(n-1)\ \cdots\ b_i(n-L+1)]^T$$

$(\cdot)^T$ denotes a vector/matrix transpose, and L is set to the length of the longest channel impulse response by assumption. Additive noise components in different channels are assumed to be uncorrelated with the source signal. The channel parameter matrix $\mathbf{H}_{\mathrm{t},i}$ is of dimension $L \times (2L-1)$ and is constructed from the channel's impulse response

$$\mathbf{h}_{\mathrm{t},i} = [h_{\mathrm{t},i,0} \quad h_{\mathrm{t},i,1} \quad \cdots \quad h_{\mathrm{t},i,L-1}]^T . \tag{3}$$

A BCI algorithm is used to estimate the channel impulse responses $\mathbf{h}_{\mathrm{t},i}$, $i = 1, 2, \cdots, M$, from the observations $\mathbf{x}_i(n)$ without utilizing any knowledge about the source signal $\mathbf{s}(n)$.

The following two assumptions (one on the channel diversity and the other on the input source signal) are made throughout this letter to guarantee an identifiable system [3].

1) The polynomials formed from $\mathbf{h}_{\mathrm{t},i}$, $i = 1, 2, \cdots, M$, are coprime, i.e., the channel transfer functions $H_{\mathrm{t},i}(z)$ do not share any common zeros.

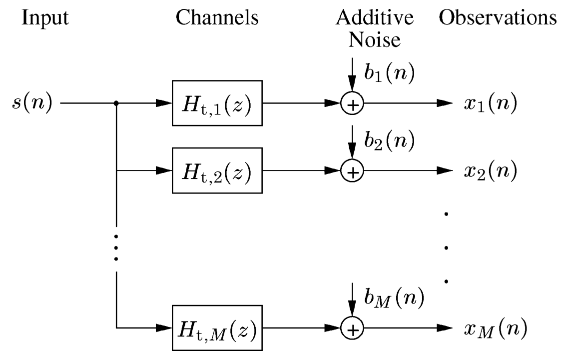


Fig. 1. Relationships between the input $s(n)$ and the observations $x_i(n)$ in a SIMO FIR system.

2) The autocorrelation matrix $\mathbf{R}_{ss} = E\{\mathbf{s}(n)\mathbf{s}^T(n)\}$ of the source signal is of full rank (such that the SIMO system can be fully excited from a perspective of system identification), where $E\{\cdot\}$ denotes mathematical expectation.

III. BCI Fundamentals and Adaptive Multichannel LMS Algorithm

For a SIMO system, the vector of channel impulse responses lies in the null space of the cross-correlation-like matrix of channel outputs [4]

$$\mathbf{R}_x \mathbf{h}_t = \mathbf{0} \tag{4}$$

where

$$\mathbf{R}_x = \begin{bmatrix} \sum_{i\neq 1}\mathbf{R}_{x_i x_i} & -\mathbf{R}_{x_2 x_1} & \cdots & -\mathbf{R}_{x_M x_1} \\ -\mathbf{R}_{x_1 x_2} & \sum_{i\neq 2}\mathbf{R}_{x_i x_i} & \cdots & -\mathbf{R}_{x_M x_2} \\ \vdots & \vdots & \ddots & \vdots \\ -\mathbf{R}_{x_1 x_M} & -\mathbf{R}_{x_2 x_M} & \cdots & \sum_{i\neq M}\mathbf{R}_{x_i x_i} \end{bmatrix}$$

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

$$\mathbf{h}_t = \left[\mathbf{h}_{t,1}^T \ \mathbf{h}_{t,2}^T \ \cdots \ \mathbf{h}_{t,M}^T\right]^T.$$

If the two conditions in Section II are met, Matrix $\mathbf{R}_x$ is rank deficient by 1 in the absence of noise and channel impulse responses and can be uniquely determined from $\mathbf{R}_x$, which contains only the second-order statistics of the system outputs. When noise is present, $\mathbf{h}_t$ would be the eigenvector of $\mathbf{R}_x$ corresponding to its smallest eigenvalue.

To develop an adaptive BCI implementation, a simple way is to take advantage of the cross relations among the outputs, as we did in an earlier study of the MCLMS algorithm [1]. By following the fact that

$$x_i * h_{t,j} = s * h_{t,i} * h_{t,j} = x_j * h_{t,i}, \quad i,j = 1,2,\cdots,M, \ i \neq j \tag{5}$$

we have, in the absence of noise, the following cross relation at time n:

$$\mathbf{x}_i^T(n)\mathbf{h}_{t,j} = \mathbf{x}_j^T(n)\mathbf{h}_{t,i}, \quad i,j = 1,2,\cdots,M, \ i \neq j. \tag{6}$$

When noise is present and/or the estimate of channel impulse responses deviates from the true value, an *a priori* error signal is produced:

$$e_{ij}(n+1) = \mathbf{x}_i^T(n+1)\mathbf{h}_j(n) - \mathbf{x}_j^T(n+1)\mathbf{h}_i(n) \qquad i,j = 1,2,\cdots,M \tag{7}$$

where $\mathbf{h}_i(n)$ is the model filter for the ith channel at time n. In order to avoid the trivial estimate of all zero elements, a unit-norm constraint is imposed on

$$\mathbf{h}(n) = \left[\mathbf{h}_1^T(n) \quad \mathbf{h}_2^T(n) \quad \cdots \quad \mathbf{h}_M^T(n)\right]^T$$

leading to the normalized error signal $\epsilon_{ij}(n+1) = e_{ij}(n+1)/\|\mathbf{h}(n)\|$. Accordingly, the cost function is formulated as

$$J(n+1) = \sum_{i=1}^{M-1}\sum_{j=i+1}^{M} \epsilon_{ij}^2(n+1) \tag{8}$$

and the update equation of the MCLMS algorithm is deduced as follows:

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

where μ is a small positive step size

$$\nabla J(n+1) = \frac{\partial J(n+1)}{\partial \mathbf{h}(n)} = \frac{2\left[\tilde{\mathbf{R}}_x(n+1)\mathbf{h}(n) - J(n+1)\mathbf{h}(n)\right]}{\|\mathbf{h}(n)\|^2} \tag{10}$$

and the rest is shown in the equation at the bottom of the page.

IV. Optimal Step Size and the Proposed Variable Step-Size MCLMS Algorithm

It was shown in [1] that the MCLMS algorithm is able to converge in the mean to the true channel impulse response vector $\mathbf{h}_t$ if the step size μ is properly specified. However, there is no guide on how to choose μ in practice. In order to avoid divergence, a conservatively small μ is usually used, which inevitably

$$\tilde{\mathbf{R}}_x(n) = \begin{bmatrix} \sum_{i\neq 1}\tilde{\mathbf{R}}_{x_i x_i}(n) & -\tilde{\mathbf{R}}_{x_2 x_1}(n) & \cdots & -\tilde{\mathbf{R}}_{x_M x_1}(n) \\ -\tilde{\mathbf{R}}_{x_1 x_2}(n) & \sum_{i\neq 2}\tilde{\mathbf{R}}_{x_i x_i}(n) & \cdots & -\tilde{\mathbf{R}}_{x_M x_2}(n) \\ \vdots & \vdots & \ddots & \vdots \\ -\tilde{\mathbf{R}}_{x_1 x_M}(n) & -\tilde{\mathbf{R}}_{x_2 x_M}(n) & \cdots & \sum_{i\neq M}\tilde{\mathbf{R}}_{x_i x_i}(n) \end{bmatrix}$$

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

sacrifices the convergence speed of the adaptive algorithm. In this section, we will show the optimal step size for the MCLMS algorithm and propose a variable step-size MCLMS algorithm.

We begin with re-examining the update equation (9). As the adaptive algorithm proceeds, the cost function $J(n+1)$ diminishes, and its gradient with respect to $\mathbf{h}(n)$ can be approximated as

$$\nabla J(n+1) \approx \frac{2\tilde{\mathbf{R}}_x(n+1)\mathbf{h}(n)}{\|\mathbf{h}(n)\|^2}. \tag{11}$$

If we remove the unit-norm constraint, a simplified UMCLMS adaptive algorithm is deduced

$$\mathbf{h}(n+1) = \mathbf{h}(n) - 2\mu\tilde{\mathbf{R}}_x(n+1)\mathbf{h}(n) \tag{12}$$

which is theoretically equivalent to the adaptive algorithm proposed in [2], although the cost functions are defined in different ways in these two adaptive BCI algorithms.

With such a simplified adaptive algorithm, the primary concern is whether it would converge to the trivial all-zero estimate. Fortunately, this will not happen, as long as the initial estimate $\mathbf{h}(0)$ is not orthogonal to the true channel impulse response vector $\mathbf{h}_\mathrm{t}$, as shown in [2]. This can be easily demonstrated by premultiplying (12) with $\mathbf{h}_\mathrm{t}^T$:

$$\mathbf{h}_\mathrm{t}^T\mathbf{h}(n+1) = \mathbf{h}_\mathrm{t}^T\mathbf{h}(n) - 2\mu\mathbf{h}_\mathrm{t}^T\tilde{\mathbf{R}}_x(n+1)\mathbf{h}(n). \tag{13}$$

Using the cross relation (6), we know $\mathbf{h}_\mathrm{t}^T\tilde{\mathbf{R}}_x(n+1) = \mathbf{0}^T$ in the absence of noise. This implies that the gradient $\nabla J(n+1)$ is orthogonal to $\mathbf{h}_\mathrm{t}$ at any time n. As a result, (13) turns out to be

$$\mathbf{h}_\mathrm{t}^T\mathbf{h}(n+1) = \mathbf{h}_\mathrm{t}^T\mathbf{h}(n). \tag{14}$$

This indicates that $\mathbf{h}_\mathrm{t}^T\mathbf{h}(n)$ is time invariant for the UMCLMS algorithm. Provided that $\mathbf{h}_\mathrm{t}^T\mathbf{h}(0) \neq 0$, $\mathbf{h}(n)$ would not converge to zero.

Decompose the model filter $\mathbf{h}(n)$ as follows:

$$\mathbf{h}(n) = \mathbf{h}_\perp(n) + \mathbf{h}_{||}(n) \tag{15}$$

where $\mathbf{h}_\perp(n)$ and $\mathbf{h}_{||}(n)$ are perpendicular and parallel to $\mathbf{h}_\mathrm{t}$, respectively. Since the gradient $\nabla J(n+1)$ is orthogonal to $\mathbf{h}_\mathrm{t}$, and $\mathbf{h}_\mathrm{t}$ is parallel to $\mathbf{h}_{||}(n)$, obviously, $\nabla J(n+1)$ is orthogonal to $\mathbf{h}_{||}(n)$ as well. Therefore, the update equation (12) of the UMCLMS algorithm can be decomposed into the following two separate equations:

$$\mathbf{h}_\perp(n+1) = \mathbf{h}_\perp(n) - \mu\nabla J(n+1) \tag{16}$$

$$\mathbf{h}_{||}(n+1) = \mathbf{h}_{||}(n). \tag{17}$$

From (16) and (17), it is clear that the UMCLMS algorithm adapts the model filter only in the direction that is perpendicular to $\mathbf{h}_\mathrm{t}$. The component $\mathbf{h}_{||}(n)$ is not altered in the process of adaptation.

As far as a general system identification algorithm is concerned, the most important performance measure apparently should be the difference between the true channel impulse response and the estimate. With a BCI method, the SIMO FIR system can be blindly identified up to a scale. Therefore, the misalignment of an estimate $\mathbf{h}(n)$ with respect to the true channel impulse response vector $\mathbf{h}_\mathrm{t}$ would be

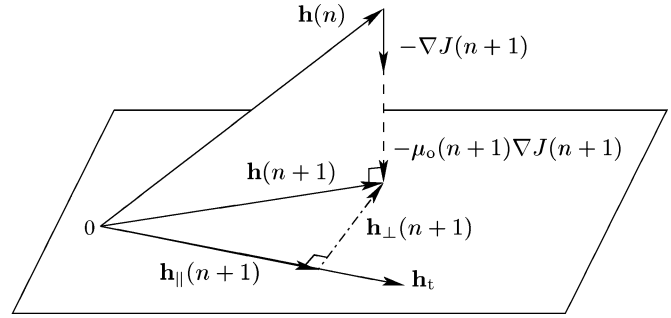


Fig. 2. Optimal step size $\mu_\mathrm{o}(n+1)$ for the unconstrained MCLMS BCI algorithm in a three-dimensional space.

$$d(n) = \min_\alpha \|\mathbf{h}_\mathrm{t} - \alpha\mathbf{h}(n)\|^2 \tag{18}$$

where α is an arbitrary scale. Substituting (15) into (18) and finding the minimum produces

$$\begin{aligned} d(n) &= \min_\alpha \left[\|\mathbf{h}(n)\|^2\alpha^2 - 2\|\mathbf{h}_{||}\|\|\mathbf{h}_\mathrm{t}\|\alpha + \|\mathbf{h}_\mathrm{t}\|^2\right] \\ &= \frac{\|\mathbf{h}_\mathrm{t}\|^2}{1+\left(\frac{\|\mathbf{h}_{||}(n)\|}{\|\mathbf{h}_\perp(n)\|}\right)^2}. \end{aligned} \tag{19}$$

Clearly, the ratio of $\|\mathbf{h}_{||}(n)\|$ over $\|\mathbf{h}_\perp(n)\|$ reflects how close the estimate is from the desired solution. With this feature in mind, the optimal step size $\mu_\mathrm{o}(n+1)$ for the UMCLMS algorithm at time $n+1$ would be the one that makes $\mathbf{h}_\perp(n+1)$ have a minimum norm, i.e.,

$$\begin{aligned} \mu_\mathrm{o}(n+1) &= \arg\min_\mu \|\mathbf{h}_\perp(n+1)\| \\ &= \arg\min_\mu \|\mathbf{h}_\perp(n) - \mu\nabla J(n+1)\|. \end{aligned} \tag{20}$$

Since $\mathbf{h}_{||}(n+1)$ is time invariant, and $\mathbf{h}_{||}(n+1)$ is orthogonal to $\mathbf{h}_\perp(n+1)$, minimizing the norm of $\mathbf{h}_\perp(n+1)$ is equivalent to minimizing the norm of $\mathbf{h}(n+1)$. As such, we have

$$\begin{aligned} \mu_\mathrm{o}(n+1) &= \arg\min_\mu \|\mathbf{h}(n+1)\| \\ &= \arg\min_\mu \|\mathbf{h}(n) - \mu\nabla J(n+1)\|. \end{aligned} \tag{21}$$

In order to minimize the norm of $\mathbf{h}(n+1) = \mathbf{h}(n) - \mu(n+1)\nabla J(n+1)$, as illustrated in Fig. 2, $\mu(n+1)$ should be chosen such that $\mathbf{h}(n+1)$ is orthogonal to $\nabla J(n+1)$. Therefore, we project $\mathbf{h}(n)$ onto $\nabla J(n+1)$ and obtain the optimal step size

$$\mu_\mathrm{o}(n+1) = \frac{\mathbf{h}^T(n)\nabla J(n+1)}{\|\nabla J(n+1)\|^2}. \tag{22}$$

Finally, this new adaptive algorithm with the optimal step size is referred to as the VSS-UMCLMS for BCI.

V. Simulations

In this section, we will evaluate the performance of the proposed VSS-UMCLMS algorithm by simulations. A comparison to the UMCLMS with a number of different prespecified step sizes is also presented.

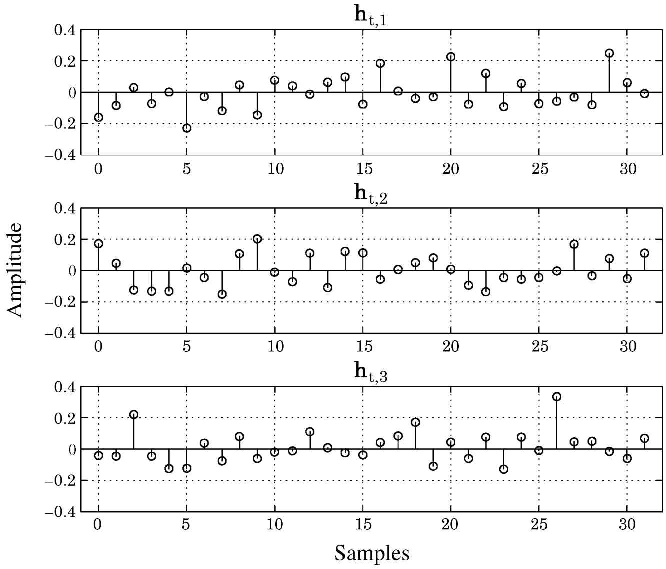


Fig. 3. Impulse responses of a single-input three-output system used in the simulation for BCI.

Similar to our earlier studies on BCI, we use the normalized projection misalignment (NPM) as a performance measure in this letter, which is given by

$$\text{NPM}(n) \triangleq \frac{\|\varepsilon(n)\|}{\|\mathbf{h}_t\|} \tag{23}$$

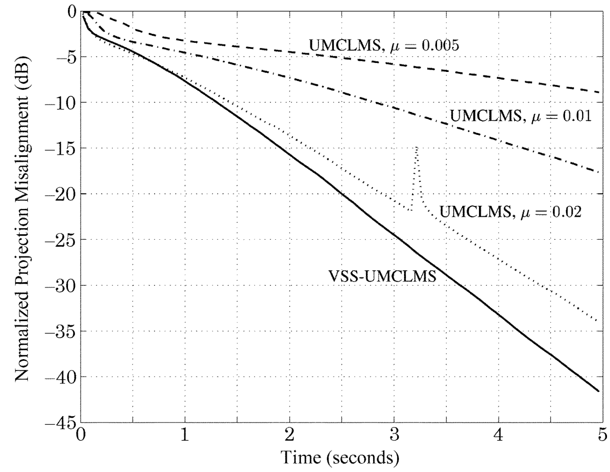
where

$$\varepsilon(n) = \mathbf{h}_t - \frac{\mathbf{h}_t^T \mathbf{h}(n)}{\mathbf{h}^T(n)\mathbf{h}(n)} \mathbf{h}(n)$$

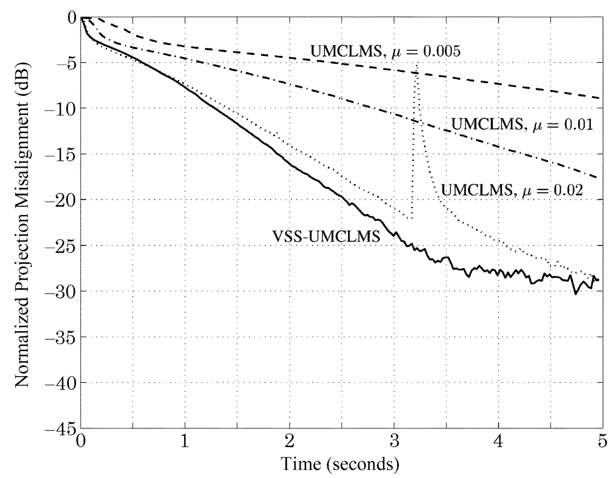
is the projection misalignment vector. By projecting $\mathbf{h}_t$ onto $\mathbf{h}(n)$ and defining a projection error, we take into account only the undesirable misalignment of the channel estimate, disregarding an arbitrary gain factor inherently associated with it [5].

The SIMO FIR system to be identified consists of $M = 3$ channels. The impulse response of each channel has $L = 32$ taps, and their coefficients are randomly generated. Fig. 3 plots these three impulse responses, which have been checked to ensure that they do not share any common zeros. The source and additive noise signals are uncorrelated, and both are white Gaussian random sequences. The sampling rate is 8 kHz. The model filter is initialized as $\mathbf{h}(0) = \mathbf{1}$ for all investigated adaptive algorithms. Note that $\mathbf{h}_t^T \mathbf{h}(0) = 0.3915 \neq 0$.

Fig. 4 shows the convergence in terms of the NPM for all the algorithms. In panel (a), noise is absent, and in panel (b), the signal-to-noise ratio (SNR) is 30 dB. Regarding the UMCLMS algorithm, a number of different step sizes were tried, and three results with $\mu = 0.02$, 0.01, and 0.005 are presented here. We see that increasing the step size would accelerate the UMCLMS algorithm to converge. However, this trend fails when μ is greater than 0.02, where the UMCLMS algorithm starts diverging. The proposed VSS-UMCLMS algorithm converges much faster than the UMCLMS algorithm, both in the absence and presence of noise. Furthermore, the final NPM for the VSS-UMCLMS algorithm is also smaller.



(a)



(b)

Fig. 4. Normalized projection misalignment of the VSS-UMCLMS and UMCLMS algorithms with three different prespecified step sizes ($\mu = 0.02$, 0.01, and 0.005) (a) in the absence of noise and (b) at 30-dB SNR.

VI. Conclusions

The optimal step size of the adaptive MCLMS algorithm for blind SIMO identification was derived, and a variable step-size unconstrained MCLMS algorithm was proposed. Compared with the conventional unconstrained MCLMS algorithm, the proposed method converges much faster and yields a more accurate estimate of the system's channel impulse responses, as demonstrated by simulations. In addition, the proposed method is much easier to use in practice, since the step size does not have to be specified in advance.

References

[1] Y. Huang and J. Benesty, "Adaptive multi-channel least mean square and Newton algorithms for blind channel identification," *Signal Process.*, vol. 82, pp. 1127–1138, Aug. 2002.

[2] H. Chen, X. Cao, and J. Zhu, "Convergence of stochastic-approximation-based algorithms for blind channel identification," *IEEE Trans. Inf. Theory*, vol. 48, no. 5, pp. 1214–1225, May 2002.

[3] 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.

[4] C. Avendano, J. Benesty, and D. R. Morgan, "A least squares component normalization approach to blind channel identification," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, vol. 4, 1999, pp. 1797–1800.

[5] 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.