14th European Signal Processing Conference (EUSIPCO 2006), Florence, Italy, September 4-8, 2006, copyright by EURASIP

IMPROVED LMS-TYPE ADAPTIVE FILTERING ALGORITHMS USING A NEW MAXIMUM EIGENVALUE BOUND ESTIMATION SCHEME

Y. Zhou, S. C. Chan, and K. L. Ho

Department of Electrical and Electronic Engineering, The University of Hong Kong
Pokfulam Road, Hong Kong
email: {yizhou, scchan, klho}@eee.hku.hk

ABSTRACT

This paper proposes a new recursive scheme for estimating the maximum eigenvalue bound for autocorrelation matrices and its application to the stepsize selection in least mean squares (LMS)-type adaptive filters. This scheme is developed from the Gershgorin circle theorem and the recursive nature of estimating the correlation matrix. The bound of the maximum eigenvalue of a $L \times L$ correlation matrix can be recursively estimated in $O(L)$ arithmetic complexity. Applying this new recursive estimate to the stepsize selection of LMS-type algorithms, the problem of over-estimating the maximum eigenvalue bound and hence the under-estimation of the stepsize in the conventional trace estimator is ameliorated. This significantly improves the transient convergence and tracking speed of LMS-type algorithms. To lower the extra steady state error caused by the use of bigger stepsizes, an effective switching mechanism is designed and incorporated into the proposed algorithms so that a smaller stepsize can be invoked near the steady state. The superior performance of the proposed algorithms is verified by numerical and computer simulations.

1. INTRODUCTION

Adaptive filters find important applications in signal processing, communications, control and many other areas [1]. Many adaptive algorithms have been proposed and one big family is the least mean squares (LMS) algorithms. The LMS-type algorithms are very widely used for their simplicity and high numerical stability. They include the LMS algorithm [2][3], the TDLMS algorithm [4][5], and the fast LMS/Newton algorithm [6], to list just a few. These algorithms have a low computational complexity of $O(L)$ (where L is the number of taps of the adaptive filter) and thus are very attractive for applications involving high-order adaptive filters. One disadvantage of the conventional LMS algorithm is its slow convergence and sensitivity to the eigenvalue spread of the input autocorrelation matrix $\boldsymbol{R}_{xx}$. The latter problem, which is usually called the eigenvalue spread problem, is alleviated in TDLMS and fast LMS/Newton algorithms by pre-whitening the input signals with appropriate transformations. Another important issue in LMS-type algorithms is the selection of an appropriate stepsize μ to ensure good performance and stability. More precisely, the mean convergence analysis of the LMS-type algorithms suggests that the stepsize should satisfy $0 < \mu < 2/\lambda_{\max}$, where $\lambda_{\max}$ is the largest eigenvalue of the input correlation matrix $\boldsymbol{R}_{xx}$ (for TDLMS, the input is the transformed and normalized input). Larger stepsizes will lead to faster initial convergence speed, but usually a higher steady-state error. Since it is computationally expensive to calculate the maximum eigenvalue, a conventional method is to estimate a bound of $\lambda_{\max}$ as $\lambda_{\max} < tr(\boldsymbol{R}_{xx})$, where $tr(.)$ is the trace operator. This is because the eigenvalues of the Hermitian matrix $\boldsymbol{R}_{xx}$ are all non-negative and $tr(\boldsymbol{R}_{xx})$ is equal to the sum of all eigenvalues of $\boldsymbol{R}_{xx}$. The bound $0 < \mu < 2/tr(\boldsymbol{R}_{xx})$ is thus widely adopted, because of the reliability and simplicity in computing $tr(\boldsymbol{R}_{xx})$, which amounts to adding the diagonal entries of $\boldsymbol{R}_{xx}$. To see that the bound is loose, we can consider a simple but commonly encountered situation where the input is white with $\boldsymbol{R}_{xx} = \boldsymbol{I}$. The maximum eigenvalue is equal to one while $tr(I)$ gives a bound of L, where $L \times L$ is the dimension of I. In other words, the stepsize will be over-estimated and the effect will become more serious if the filter length is increased. This explains the slow convergence of LMS-type algorithms when the filter length is increased and the speedup observed in some LMS-type algorithms when the filter is implemented in cascade.

In this paper, we propose a new recursive scheme for estimating the maximum eigenvalue bound for autocorrelation matrices. The algorithm is developed from the Gershgorin circle theorem (GCT) [7] and the recursive nature of estimating the correlation matrix. It requires only $O(L)$ arithmetic complexity. Simulation results show that the proposed estimator significantly improves the convergence and tracking speed of the LMS-type algorithms. The resultant algorithms using this scheme are referred to as the Recursive Maximum Eigenvalue bound-based (RME) LMS-type algorithms. On the other hand, due to consideration of misadjustment, the stepsize is usually chosen as $\mu \approx M/L$ where M is the value of intended misadjustment. Therefore, in order to avoid excessive steady state error, we also propose an effec-

tive switching mechanism based on the work in [8] so that a smaller stepsize can be invoked near the steady state. The superior performance of the proposed algorithms is verified by numerical and computer simulations. This paper is organized as follows: some conventional LMS-type algorithms are briefly reviewed in section 2. The new recursive maximum eigenvalue bound estimation scheme and its application to these LMS-type algorithms are presented in section 3. Experimental results and comparisons are presented in section 4. Conclusions are drawn in section 5. To keep the presentation simple, all the derivations are given for real-valued signals.

2. CONVENTIONAL LMS-TYPE ALGORITHMS

Consider the identification of an unknown system with impulse response $\boldsymbol{W}^*$ shown in Fig. 1. The unknown system and the adaptive filter with impulse response $\boldsymbol{W}(n)=\begin{bmatrix}w_1(n) & w_2(n) & \cdots & w_L(n)\end{bmatrix}^T$ are simultaneously excited by an input signal $x(n)$. The adaptive filter continuously adjusts its weight coefficients according to certain algorithms to minimize certain performance criterion such as the mean-square-error (MSE) of the instantaneous estimation error $e(n)$, which is the difference between the desired signal $d(n)$ and the filter output $y(n)$. $d_0(n)$ is the output of the unknown system and $\eta_0(n)$ represents any possible modeling error and/or background noises. Some LMS-type algorithms are summarized as follows:

2.1. LMS algorithm

LMS algorithm is obtained by using the instantaneous estimate of the gradient $\hat{\nabla}J(\boldsymbol{W})=-2e(n)\boldsymbol{X}(n)$ in place of the true gradient in the steepest-descent algorithm. It can be summarized as

$$d(n)=\boldsymbol{W}^{*T}\boldsymbol{X}(n)+\eta_0(n), \tag{1}$$

$$e(n)=d(n)-\boldsymbol{W}^T(n-1)\boldsymbol{X}(n), \tag{2}$$

$$\boldsymbol{W}(n)=\boldsymbol{W}(n-1)+\mu_{LMS}e(n)\boldsymbol{X}(n). \tag{3}$$

where $\boldsymbol{X}(n)=\begin{bmatrix}x_1(n) & x_2(n) & \cdots & x_L(n)\end{bmatrix}^T$, μ_{LMS} is the constant stepsize controlling the convergence speed and steady state error.

2.2. TDLMS algorithm

The difference between TDLMS and LMS lies in the orthogonal transformation of the original input vector prior to the filtering process and the subsequent normalization in the updating algorithm, which can be summarized as

$$\boldsymbol{S}(n)=\boldsymbol{T}\boldsymbol{X}(n), \tag{4}$$

$$\boldsymbol{W}(n)=\boldsymbol{W}(n-1)+\mu_{TDLMS}e(n)\Lambda^{-2}(n)\boldsymbol{S}(n). \tag{5}$$

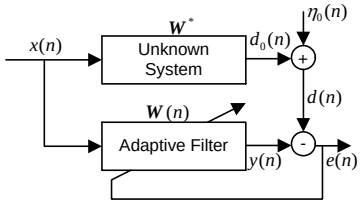


Figure 1 – System identification structure

where the matrix $\boldsymbol{T}$ represents an orthogonal transformation such as the discrete cosine transform (DCT) and discrete Fourier transform (DFT).

$\Lambda^{-2}(n)=diag(\sigma_{x_1}^{-2}(n)\quad \sigma_{x_2}^{-2}(n)\quad \cdots\quad \sigma_{x_L}^{-2}(n))$,

where $\sigma_{x_i}^2(n)$, $i=1,2,\cdots,L$, is the estimated input power on the i-th tap which can be updated as $\sigma_{x_i}^2(n)=\lambda_\sigma\sigma_{x_i}^2(n-1)+(1-\lambda_\sigma)x_i^2(n)$, λ_σ is a positive constant forgetting factor close to but less than 1.

2.3. Fast LMS/Newton algorithm

In the LMS/Newton algorithm, $\hat{\boldsymbol{R}}_{xx}^{-1}(n)$, the estimate of the inverse of the input autocorrelation matrix $\boldsymbol{R}_{xx}(n)$, is employed in the weight update equation to decorrelate the input vector so as to achieve faster convergence. In [6], the following fast LMS/Newton algorithm was proposed:

$$e(n)=d(n-M)-\boldsymbol{X}^T(n-M)\boldsymbol{W}(n), \tag{6}$$

$$\boldsymbol{u}_a(n)=\boldsymbol{L}_2(n)\tilde{\boldsymbol{D}}^{-1}(n)\boldsymbol{L}_1(n)\boldsymbol{X}_E(n), \tag{7}$$

$$\boldsymbol{W}(n+1)=\boldsymbol{W}(n)+\mu_{fast\ LN}e(n)\boldsymbol{u}_a(n). \tag{8}$$

The criteria of selecting μ for these three LMS-type algorithms comply with those introduced in section 1, that is,

$$0<\mu_{LMS}<K/tr(\boldsymbol{R}_{xx}), \tag{9}$$

$$\mu_{TDLMS},\mu_{fast\ LN}\approx M/L. \tag{10}$$

where K is a positive constant that can be flexibly selected.

3. PROPOSED RECURSIVE MAXIMUM EIGENVALUE BOUND ESTIMATION SCHEME AND ITS APPLICATION TO LMS-TYPE ALGORITHMS

To avoid over-estimating the bound of μ as in (9) and (10), we shall seek for a more accurate scheme for recursively updating $\lambda_{\max}(n)$. The Gershgorin circle theorem (GCT) identifies a region in the complex plane that contains all the eigenvalues of a square matrix. More specifically, for an $L\times L$ matrix $\boldsymbol{R}$, define $R_i=\sum_{\substack{j=1\\ j\neq i}}^{L}|r_{ij}|$, $i=1,2,\cdots,L$, where r_{ij} is the (i,j)-*th* element of $\boldsymbol{R}$. Then each eigenvalue of $\boldsymbol{R}$ is in at least one of the disks

$$\{\lambda:|\lambda-r_{ii}|<R_i\}. \tag{11}$$

In adaptive filtering, $\boldsymbol{R}$ is the input autocorrelation matrix $\boldsymbol{R}_{xx}$, which needs to be estimated recursively. A simple way is to update $\boldsymbol{R}_{xx}(n)$ at each time instant, which usually needs an $O(L^2)$ arithmetic complexity. Next we propose a new maximum eigenvalue bound estimation scheme based on GCT.

Let $B_i(n)$ be the recursive estimate of $R_i(n)$, Then

$$\begin{aligned} B_i(n) &= \lambda_B B_i(n-1) + (1-\lambda_B)\cdot\sum_{\substack{j=1\\ j\neq i}}^{L} |r_{ij}(n)| \\ &= \lambda_B B_i(n-1) + (1-\lambda_B)\cdot\sum_{\substack{j=1\\ j\neq i}}^{L} |x_i(n)x_j(n)| \\ &= \lambda_B B_i(n-1) + (1-\lambda_B)\cdot[|x_i(n)|\cdot(\sum_{j=1}^{L}|x_j(n)| - |x_i(n)|)] \\ &= \lambda_B B_i(n-1) + (1-\lambda_B)\cdot[|x_i(n)|\cdot(s(n) - |x_i(n)|)] \end{aligned} \tag{12}$$

where $s(n)=\sum_{j=1}^{L}|x_j(n)|$, λ_B is also a positive constant forgetting factor close to but less than 1. Since the eigenvalue of the autocorrelation matrix is positive, a simpler bound for the maximum eigenvalue is

$$\lambda_{\max}(n) \le \lambda^B_{\max}(n) = \max_i(B_i(n) + \sigma^2_{x_i}(n)). \tag{13}$$

The computational complexity is $O(L)$ ($3L$+2 additions, L multiplications for FIR filter case and $4L$-1 additions, L multiplications for array case). And if further complexity reduction is needed, we can use the following approximation

$$B(n) = \lambda_B B(n-1) + (1-\lambda_B)\cdot[|x_{\max}(n)|\cdot s(n) - |x_{\min}(n)|^2], \tag{14}$$

$$\lambda_{\max}(n) \le \lambda^B_{\max}(n) = B(n) + \sigma^2_{x_{\max}}(n). \tag{15}$$

Compare (12) ~ (15) with the conventional updating approach of $\boldsymbol{R}_{xx}(n)$, where the instantaneous estimate for the autocorrelation matrix is $\boldsymbol{X}(n)\boldsymbol{X}^T(n)$ and the recursive estimate of $\boldsymbol{R}_{xx}(n)$ is $\lambda_B\boldsymbol{R}_{xx}(n-1)+\boldsymbol{X}(n)\boldsymbol{X}^T(n)$. This reveals the maximum eigenvalue bound estimated by the proposed scheme is bigger than that obtained directly from the GCT. Nevertheless, the new estimate predicted by the proposed scheme is still smaller than the traditional estimate $tr(\boldsymbol{R}_{xx})$, which can be illustrated in the experiments in section 4. This alleviates the over-estimation problem at the expense of moderate extra computation. Next we shall apply this new scheme in the LMS-type algorithms introduced in section 2. Define

$$\mu_{RME}(n) = K'/\lambda^B_{\max}(n). \tag{16}$$

we can then replace μ_{LMS}, μ_{TDLMS} and $\mu_{fast\ LN}$ in (3), (5) and (8) respectively with (16). Since it takes several iterations for $\lambda^B_{\max}(n)$ to stabilize, $\mu_{RME}(n)$ is chosen as the above constant stepsizes for the first S time instants. The value of S can be experimentally decided. The resultant algorithms are called the RME-LMS 1, RME-TDLMS 1 and RME-fast LMS/Newton 1algorithms.

Although these algorithms possess improved initial convergence speed, the steady state error will increase due to the larger stepsizes used. To overcome this problem, we proposed to switch $\mu_{MEE}(n)$ back to μ_{LMS}, μ_{TDLMS} and $\mu_{fast\ LN}$ when the algorithms are about to approach their steady states. We now propose a measure based on the work in [8]. This so-called GP-APA algorithm motivated us to measure the convergence status of the algorithm through the approximated derivatives of the filter weight as follows:

$$\|\hat{\boldsymbol{c}}(n-1)\|_1 = \sum_{i=1}^{L}|\hat{c}_i(n-1)|, \tag{17}$$

$$\hat{c}_i(n) = w_i(n-1) - \hat{w}_i(n-1),$$

$$\hat{w}_i(n) = \eta\hat{w}_i(n-1) + (1-\eta)w_i(n-1), \quad i=1,2,\cdots,L. \tag{18}$$

where $\hat{w}_i(n)$ is the averaged tap coefficient and $\hat{c}_i(n)$ is the averaged estimate of the coefficient gradient. $\|\cdot\|_1$ represents the l_1 norm of a vector and η is the constant forgetting factor. The value of $\|\hat{\boldsymbol{c}}(n-1)\|_1$ will decrease and converge gradually from its initial value to a very small value when the algorithm is about to converge to its steady state. Therefore, there always exists a threshold below which we can say that the algorithm is near convergence. We now propose a method to compute this threshold for making the switching decision. More precisely, we first compute the absolute value of the approximate derivative of $\|\hat{\boldsymbol{c}}(n-1)\|_1$ as $G_c(n) = |\|\hat{\boldsymbol{c}}(n)\|_1 - \|\hat{\boldsymbol{c}}(n-1)\|_1|$ and then the decaying ratio $\chi(n) = G_c(n)/G_c^0$, where G_c^0 represents the approximation of the initial level of $G_c(n)$, which is obtained by averaging $G_c(k)$ from $k=1,2,\cdots P$. It provides a reference to measure the decay of the coefficients weight vector. A small value of $\chi(n)$ indicates a diminished variations in the weight vector, and the filter is likely to be near the end of its initial converging period. By choosing appropriately a threshold, say $\hat{\chi}$, it is possible to compare $\chi(n)$ against this threshold to determine whether switching is necessary. When $\chi(n)$ is larger than $\hat{\chi}$, the algorithm is likely to be in its initial convergence stage and $\mu_{MEE}(n)$ should be used to achieve fast convergence speed. On the other hand, when $\chi(n)$ falls below $\hat{\chi}$, the algorithm is likely to converge to its steady state and constant stepsizes should be invoked to further lower the steady state error. To guarantee that $\chi(n)$ has actually decreased below the threshold, the switching decision should be made if $\chi(n)$ is less than $\hat{\chi}$ for Q consecutive observations, where Q denotes the decision window length. The parameters $\hat{\chi}$, P and Q can be chosen experimentally in practical applications and simulation results show that the performance of the proposed algorithms is not very sensitive to these values if they are reasonably chosen. The

resultant algorithms after incorporating the switching mechanism are called RME-LMS 2, RME-TDLMS 2 and RME-fast LMS/Newton 2 algorithms to differentiate them from those introduced previously without the switching mechanism.

4. SIMULATION RESULTS

We now evaluate the performance of the proposed algorithms by computer simulation of the system identification problem depicted in Fig. 1. The order of the unknown system is set to 50 and the coefficients are randomly generated and normalized to unit power. The power of the additive white Gaussian noise $\eta_0(n)$ is set to be 0.0001. Two kinds of input signals are employed: a white Gaussian input with zero mean, unit variance and $tr(\boldsymbol{R}_{xx}) = L$, and a colored input derived from a first-order AR process driven by a unit power white sequence with coefficient $a = 0.7$, $tr(\boldsymbol{R}_{xx}) = \sigma_I^2 L/(1-a^2)$. For the RME algorithms, $\lambda_B = 0.95$, $K' = 2/3$, S=50, K in (9) are selected to be 1/10 and 1/3 for the purpose of comparison and $M = 10\%$ in (10). For the switching mechanism, $\hat{\chi} = 0.1$, $P = 20$, and $Q = 100$. The proposed RME-LMS, RME-TDLMS, RME-fast LMS/Newton algorithms and their conventional counterparts are tested in four experiments. All the results are averaged over 100 independent runs. **Exp. 1** RME-LMS vs. LMS. Fig. 2 (a) and Fig. 3 (a) show the results of using white and colored inputs respectively. The RME-LMS algorithms exhibit significantly improved transient convergence speed and algorithm 2 has very low steady state error. Fig. 2 (b) and Fig. 3 (b) illustrate the curves of $\lambda_{\max}^B(n)$ and $\lambda_{\max}(n)$ computed from $\boldsymbol{R}_{xx}(n)$. **Exp. 2** Test tracking ability. Colored input is employed. The slowly varying system is $w_i(n+1) = w_i(n) + \varepsilon|w_i(n)|v_i(n)$, $i = 1,2,\cdots,L$, where ε is a small constant equal to 0.02 and $v_i(n)$'s are a set of independent Gaussian white noise sequences with unit variance. The performance index is the sum of squared coefficient error (MSD). Fig. 4 reveals that RME-LMS 1 and 2 algorithms have the same behavior and outperform the LMS algorithm. **Exp. 3** RME-TDLMS vs. TDLMS. Colored input is employed. Similar observations to experiment 1 were made from Fig. 5 that the RME-TDLMS algorithms outperform their conventional counterpart in transient convergence speed and algorithm 2 can also provide low steady state error. **Exp. 4** RME-fast LMS/Newton vs. fast LMS/Newton. The input signal $x(n)$ is modeled as a 5-th order AR process with coefficients [1 -0.65 0.693 -0.22 0.309 -0.177] as given in [6]. From Fig. 6 it can be seen that the RME-fast LMS/Newton algorithms have similar convergence speed to the conventional algorithm when $K = 1/3$. However, the former still converge much faster than the latter when $K = 1/10$.

5. CONCLUSION

A new recursive maximum eigenvalue bound estimation scheme for autocorrelation matrices based on the Gershgorin circle theorem is presented in this paper. It has a computational complexity of $O(L)$. By applying this new bound estimate scheme to conventional LMS-type algorithms, faster transient convergence and tracking speed are obtained. To avoid the extra steady state error caused by the use of larger stepsizes, an effective switching mechanism is designed and incorporated into the proposed algorithms so that a sufficiently small constant stepsize can be invoked in the steady state to reduce the extra error.

REFERENCES

[1] S. Haykin, *Adaptive Filter Theory*, 4th ed. Prentice Hall, Englewood Cliffs, NJ, 2001.

[2] B. Widrow, J. McCool, M. G. Larimore, and C. R. Johnson, Jr., "Stationary and nonstationary learning characteristics of the LMS adaptive filter," *Proc. IEEE*, vol. 64, pp. 1151-1162, Aug. 1976.

[3] A. Feuer and E. Weinstein, "Convergence analysis of LMS filters with uncorrelated Gaussian data," *IEEE Trans. On Acoustics, Speech, and Signal Processing*, vol. 33, no. 1, pp. 222-230, Feb. 1985.

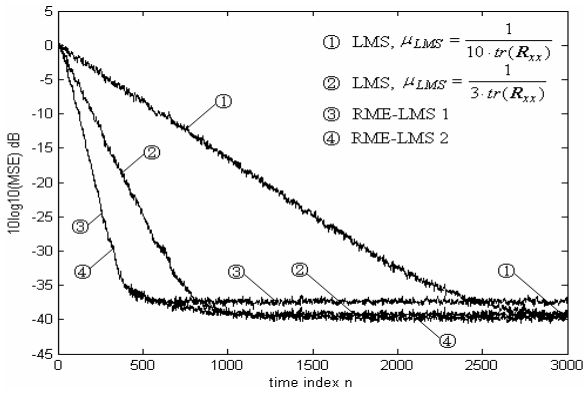
[4] S. S. Narayan, A. M. Peterson and M. J. Narasimha, "Transform domain LMS algorithm," *IEEE Trans. On Acoustic, Speech and Signal Processing*, vol. 31, no. 3, pp. 609-615, Jun. 1983.

[5] F. Beaufays, "Transform-domain adaptive filters: an analytical approach," *IEEE Trans. On Signal Processing*, vol. 43, no. 2, pp. 422-431, Feb. 1995.

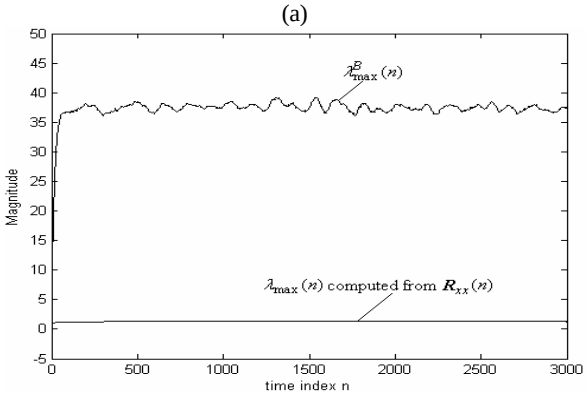
[6] B. F. Boroujeny, "Fast LMS/Newton algorithms based on autoregressive modelling and their application to acoustic echo cancellation," *IEEE Trans. On Signal Processing*, vol. 45, pp. 1987-2000, Aug. 1997.

[7] B. Noble and J. W. Daniel, *Applied Linear Algebra*, 3rd ed. Prentice Hall, Englewood Cliffs, NJ, 1988.

[8] O. Hoshuyama, R. A. Goubran and A. Sugiyama, "A generalized proportionate variable step-size algorithm for fast changing acoustic environments," *IEEE ICASSP-04*, vol. 4, pp. 161-164, May 2004.

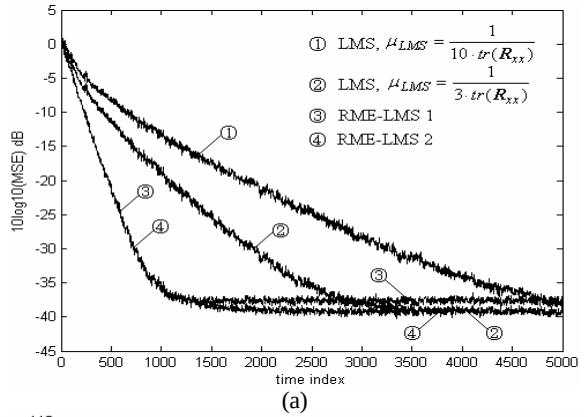


(a)

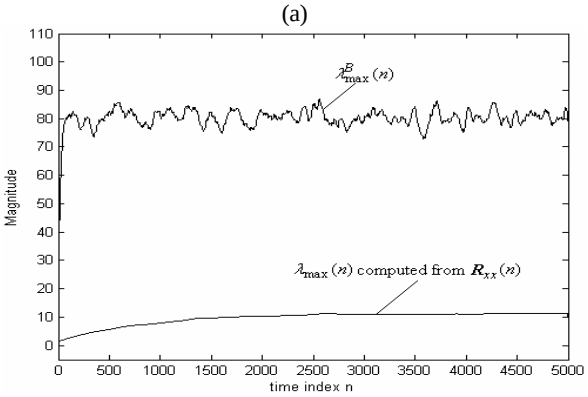


(b)

Figure 2 –Example of RME-LMS vs. LMS: white input case (a). MSE vs. time *n*; (b) $\lambda^{B}_{\max}(n)$.



(a)



(b)

Figure 3 –Example of RME-LMS vs. LMS: colored input case (a). MSE vs. time *n*; (b) $\lambda^{B}_{\max}(n)$.

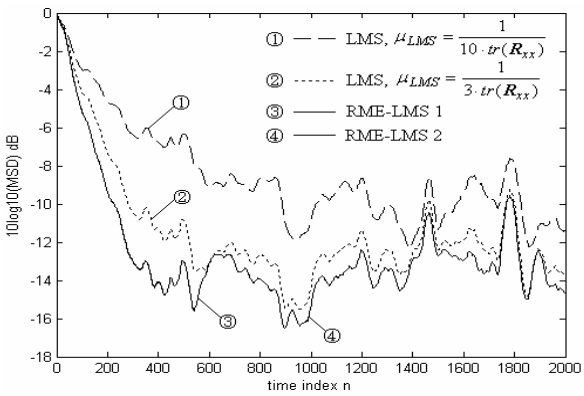


Figure 4 –Example of tracking slowly varying system parameters. RME-LMS vs. LMS: colored input case. (MSD results vs. time *n*)

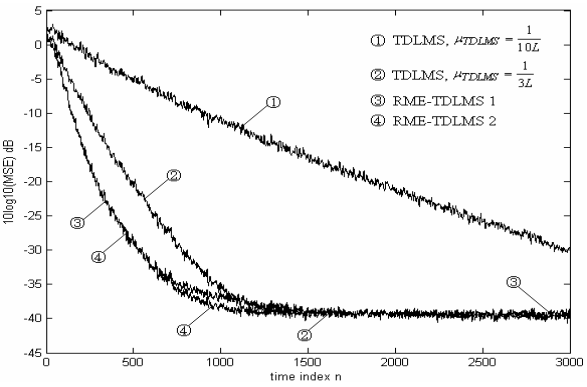


Figure 5 –Example of RME-TDLMS vs. TDLMS: colored input case.

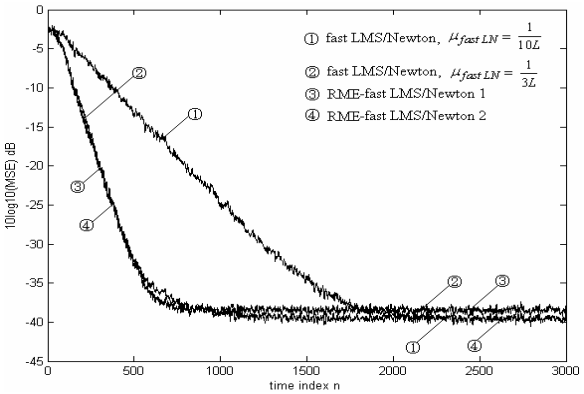


Figure 6 –Example of RME-fast LMS/Newton vs. fast LMS/Newton: colored input case.