LMS Algorithm with an Adaptive Neural Network Cost Function

JENG-KUANG HWANG *, YUAN-PING LI
* Department of Communication Engineering
Department of Electrical Engineering
Yuan-Ze University
135, Yuan-Tung Rd., Chungli City 32026
TAIWAN
eejhwang@saturn.yzu.edu.tw; s909107@mail.yzu.edu.tw

Abstract: - We propose a new LMS algorithm with an adaptive neural network cost function (ANNCFLMS) for application to unknown channel estimation or system identification. The algorithm employs the weighted average of a neural network with two input signals—the squared errors at adjacent time intervals—to modify the cost function and update the respective weight according to a gradient descent algorithm designed to track the minimum mean squared error (MSE). For fast convergence, the step-size updates recursively until the modified cost function attains its minimum value. Simulation results demonstrate that the proposed algorithm converges faster and is especially robust in low-SNR or colored input environments.

Key-Words: - Adaptive filters, neural network, least mean square (LMS), variable step-size, channel estimation

1 Introduction

The least mean square (LMS) algorithm is the most popular and widely used due to its simplicity and robustness for adaptive finite-impulse response (FIR) filters. It is described by the following [1]:

$$e(n) = d(n) - \mathbf{w}^T(n)\mathbf{x}(n) \quad (1)$$

$$\mathbf{w}(n+1) = \mathbf{w}(n) + \mu e(n)\mathbf{x}(n) \quad (2)$$

where $e(n)$ is the instantaneous error at the output of the filter for the time instant n, $d(n)$ is the desired signal, $\mathbf{x}(n) = [x(n),\ldots, x(n-N+1)]^T$ is the input signal vector, N is the length of the filter, $(\cdot)^T$ is the vector transpose operator, and $\mathbf{w}(n) = [w_0(n),\ldots, w_{N-1}(n)]^T$ is the filter coefficient (weight) vector. The step-size parameter μ is critical to the performance of the LMS and determines how fast the algorithm converges along the error performance surface defined by the cost function $(1/2)e^2(n)$. As is well known, if the step-size is large, the convergence rate of the LMS algorithm will be rapid, but the steady-state MSE will increase. In contrast, if the step-size is small, the steady-state MSE will be small, but the convergence rate will be slow. Thus, the step-size provides a tradeoff between the convergence rate and the size of the steady-state MSE of the LMS algorithm.

Many variable step-size LMS (VSLMS) algorithms have been proposed that improve on the performance of the LMS algorithm by using large step-sizes at the early stages of the adaptive process and small step-sizes when the system approaches convergence. Typical methods can be found in [2–9]. The mathematical formulae for updating the step-size $\mu(n)$ in these algorithms are summarized in Table 1, based upon the equations describing the LMS algorithm,

In practical applications, high-level measurement noise or colored input data tends to deteriorate the convergence performance and, in order to overcome this problem, the step-size in [3] is controlled by the squared instantaneous error, and robustness to uncorrelated noise is introduced in [5] by using the squared autocorrelation of errors at adjacent intervals. The step-size in [4] is adjusted by the inner product between adjacent gradient vectors and this is improved upon in [6] (a simplified version of the original algorithms in [2,4]), where a smoothing operation is exploited on one gradient vector to reduce the influence of measurement noise. The step-size in [7] is controlled by exploiting the augmented Lagrangian multipliers and knowledge of the channel noise variance to constrain the MSE minimized to the noise variance. This is expanded in [9] by using the same constrained method by which the noise variance is estimated (i.e., without any assumptions of noise required). The learning rate in [8] represents an extension of the normalized LMS (NLMS) algorithm by means of an additional gradient adaptive term in the denominator of the learning

Table 1. Simulation parameters of the adaptive algorithms for the channel estimation problem and summary of the step-size updates of their algorithms

Algorithm	Parameters	SNR = 0 dB	SNR = 10 dB	Update of the step-size
LMS	μ	0.004	0.004	A fixed value μ
GASS [4]	ρ	6.7×10^{-6}	6.7×10^{-5}	$\mu(n)=\mu(n-1)+\rho e(n)e(n-1)\mathbf{x}^T(n)\mathbf{x}(n-1)$
MVSS [5]	$\alpha, \beta,$ γ	0.97, 0.99, 0.0133	0.97, 0.99, 0.133	$p(n)=\beta p(n-1)+(1-\beta)e(n)e(n-1),$ $\mu(n+1)=\alpha\mu(n)+\gamma p^2(n)$
VSLMS [6]	β, ρ	0.95, 2×10^{-5}	0.95, 2×10^{-4}	$\mathbf{p}(n)=\beta\mathbf{p}(n-1)+e(n-1)\mathbf{x}(n-1),$ $\mu(n)=\mu(n-1)+\rho e(n)\mathbf{x}^T(n)\mathbf{p}(n)$
NCLMS [7]	$\mu_\alpha, \mu_\lambda,$ γ	0.0027, 0.01, 14.1	0.0027, 0.01, 141.07	$\lambda(n+1)=\lambda(n)+\mu_\lambda(\frac{1}{2}(e^2(n)-\sigma_n^2)-\lambda(n)),$ $\mu(n+1)=\mu_\alpha(1+\gamma\lambda(n))$
GNGD [8]	μ, ρ	0.015, 0.002	0.015, 0.002	$\varepsilon(n)=\varepsilon(n-1)-\rho\mu\frac{e(n)e(n-1)\mathbf{x}^T(n)\mathbf{x}(n-1)}{(\lVert\mathbf{x}(n-1)\rVert_2^2+\varepsilon(n-1))^2},$ $\eta(n)=\frac{\mu}{\lVert\mathbf{x}(n)\rVert_2^2+\varepsilon(n)}$
AECLMS [9]	$\mu_\alpha,$ $\mu_\lambda, \gamma,$ μ_ζ, ρ	0.0022, 6×10^{-4}, 15.6, 6×10^{-4}, 0.25	0.0022, 6×10^{-5}, 156, 6×10^{-5}, 2.5	$\zeta(n+1)=\zeta(n)+\mu_\zeta(\gamma\lambda(n)-2\rho\zeta(n)),$ $\lambda(n+1)=\lambda(n)+\mu_\lambda(\frac{1}{2}(e^2(n)-\zeta(n))-\lambda(n)),$ $\mu(n+1)=\mu_\alpha(1+\gamma\lambda(n))$
ANNCFLMS	$\alpha,$ $\rho, \mu(0)$	0.00025, 4×10^{-5}, 0.013	0.002, 3×10^{-4}, 0.025	$\lambda(n+1)=\lambda(n)+\rho\varepsilon(n)\Phi'(J(n))\left[\frac{e^2(n-1)-e^2(n)}{(1+\lambda(n))^2}\right]$ $\mu(n)=\mu(n-1)+\frac{2\alpha}{1+\lambda(n)}\Phi'(J(n))\times\left[e(n)e(n-1)\mathbf{x}^T(n)\mathbf{x}(n-1)+\lambda(n)\times e(n-1)e(n-2)\mathbf{x}^T(n-1)\mathbf{x}(n-2)\right]$

Note that parameters represented by the same symbols in different algorithms are not necessarily related.

rate of the NLMS; this preserves stability for close-to-zero input vectors. (See Table 1)

This paper proposes an approach in which the cost function of the weighted-averaging of a neural network is used to enhance the convergence rate and obtain a low steady-state misadjustment error than the existing schemes for application, such as unknown channel estimation.

2 ANNCFLMS Algorithm

Consider a neural network (NN) with a weighted average net function having two input taps and employing a threshold activation function having one output, as shown in Fig. 1. The net function $J(n)$ is defined by

$$J(n)=\frac{e^2(n)+\lambda(n)e^2(n-1)}{1+\lambda(n)}. \tag{3}$$

The output $J_\Phi(n)$ is obtained from $J(n)$ via the activation function $\Phi(J(n))$:

$$J_\Phi(n)=\Phi\left(J(n)\right)=\frac{1-e^{-J(n)}}{1+e^{-J(n)}}, \tag{4}$$

where $e^2(n)$ and $e^2(n-1)$ (squared errors at time index n and $n-1$ from (1)) are the input signals, and the weights for the NN are $1/(1+\lambda(n))$ and $\lambda(n)/(1+\lambda(n))$. Consider that the cost function of the ANNCFLMS for online learning is defined by

$$E(n)=\frac{1}{2}\varepsilon^2(n), \tag{5}$$

$$\begin{aligned}\varepsilon(n)&=J_{\min}-J_\Phi(n)\\&=\sigma_d^2-\mathbf{p}^T\mathbf{R}^{-1}\mathbf{p}-J_\Phi(n),\end{aligned} \tag{6}$$

where $J_{\min}$ denotes the minimum MSE produced by the Wiener filter, $\sigma_d^2=E\{d^2(n)\}$ is the variance of the desired response $d(n)$, $\mathbf{R}=E\{\mathbf{x}(n)\mathbf{x}^T(n)\}$ denotes the correlation matrix of the tap input $\mathbf{x}(n)$, and $\mathbf{p} = E\{\mathbf{x}(n)d(n)\}$ denotes the cross-correlation vector between the tap inputs of the filter and the desired response.

The concept involves using the weighted average sum of an NN with two input signals—the squared errors at time index n and $n-1$—to modify the cost function and update the respective weight according to the gradient descent algorithm to track $J_{\min}$. The result leads to the gradient $\nabla_{\lambda(n)}E(n)$ being minimized. The cost function (5) is minimized with regard to $\lambda(n)$ as follows:

$$\begin{aligned}\lambda(n+1) &= \lambda(n) - \rho\nabla_{\lambda(n)}E(n) \\ &= \lambda(n) + \rho\varepsilon(n)\Phi'\big(J(n)\big) \\ &\quad\times\left[\frac{e^2(n-1)-e^2(n)}{\big(1+\lambda(n)\big)^2}\right], \end{aligned} \tag{7}$$

where ρ is learning rate parameter and $\Phi'(J(n)) = 2e^{-J(n)}/(1+e^{-J(n)})^2$ denotes the first derivative of $\Phi(J(n))$. This implies that $J_\Phi(n)$ can be minimized when $J_{\min}$ is tracked by $J_\Phi(n)$. Therefore, (4) is minimized according to the gradient descent algorithm with regard to the step-size $\mu(n)$ as follows:

$$\begin{aligned}\mu(n) &= \mu(n-1) - \alpha\nabla_{\mu(n-1)}J_\Phi(n) \\ &= \mu(n-1) + \frac{2\alpha}{1+\lambda(n)}\Phi'\big(J(n)\big) \\ &\quad\times\big[e(n)e(n-1)\mathbf{x}^T(n)\mathbf{x}(n-1) + \lambda(n) \\ &\quad\times e(n-1)e(n-2)\mathbf{x}^T(n-1)\mathbf{x}(n-2)\big] \end{aligned} \tag{8}$$

, where α is learning rate parameter. The update equation for the adaptive filter weights of the proposed algorithm can be rewritten as follows:

$$\mathbf{w}(n+1) = \mathbf{w}(n) + \mu(n)e(n)\mathbf{x}(n). \tag{9}$$

Hence, (1) and (7)–(9) constitute the proposed ANNCFLMS algorithm.

2.1 Convergence Analysis and Computational Complexity of ANNCFLMS Algorithm

The classical analysis of ANNCFLMS in terms of convergence in the mean, mean-squared, and steady-state follows the well-known analyses from the literature [4,8,10,11]. The adaptive step-size of ANNCFLMS is essentially bounded by the stability limits of the step-size of the LMS algorithm. (A sufficient, but not necessary, condition on $\mu(n)$ to ensure mean-squared convergence of the adaptive

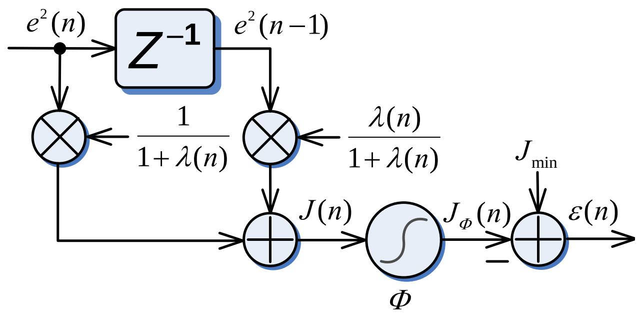


Fig. 1. The ANNCFLMS algorithm modifies the cost function using an NN architecture to track the minimum MSE.

filter is [3–5,7,9]: $0 < \mu < 2/(3\text{tr}\{\mathbf{R}\})$, where $\text{tr}\{\cdot\}$ denotes the trace of the matrix.)

The computational complexity of ANNCFLMS may appear to be somewhat high when compared with that of other adaptive algorithms, as observed in Table 1. But it is possible to reduce the computational complexity of ANNCFLMS as follows: 1) use a simple look-up table method or a lower-order Taylor series expansion to obtain an approximation of the sigmoid nonlinear activation function for the NN; 2) exploit a known channel noise variance instead of the tracked minimum MSE ($J_{\min}$); and 3) stop adaptation after convergence. In the experiments, however, ANNCFLMS was found suitable for general environments and was especially robust in low-SNR or colored input environments due to the nonlinear nature of the neural network. The results of the simulation are shown in the next section.

3 Simulation Results

Consider a FIR channel estimation problem and the following channel (as used in [7,9]):

$$[0.227, 0.460, 0.688, 0.460, 0.227]^T.$$

The input vectors are white Gaussian sequences with zero mean and the autocorrelation matrix $\mathbf{R} = \mathbf{I}$. The SNR is calculated by SNR = $10\log_{10}(E[y^2(n)/E[v^2(n)])$, where $y(n)$ is the output of the unknown system, and $v(k)$ is the system noise, $y(n) = \mathbf{w}_o^T\mathbf{x}(n)$, $\mathbf{w}_o$ is the optimal tap-weight vector. The measurement noise $v(n)$ is added to $y(n)$ so that SNR = 0 and 10 dB. The simulation parameters used here and in other algorithms are listed in Table 1 and set to achieve a steady-state misadjustment of 0.01 according to the data in the original papers. The results presented in this paper are averages taken over 1000 independent runs.

3.1 White Gaussian Input

Fig. 2 compares the convergence curves of the MSE for the LMS, GASS in [4], MVSS in [5], VSLMS in [6], NCLMS in [7], GNGD in [8], AECLMS in [9], and proposed ANNCFLMS algorithm when the noise variance is 1.0 (SNR = 0 dB) and 0.1 (SNR = 10 dB). The experimental misadjustments of the algorithm are shown in Table 2. As can be seen, the convergence of the MSE for the proposed algorithm is fastest than that of the other algorithms at low SNR (0 dB) and slightly faster than those of the NCLMS and AECLMS algorithms at high SNR (10 dB).

3.2 Nonwhite Gaussian Input

Consider the experiment adopted in [4,7]: the input vectors $\{\mathbf{x}(n)\}$ are obtained from the output of a third-order low-pass filter with the transfer function

$$H(z) = \frac{0.44}{1-1.5z^{-1}+z^{-2}-0.25z^{-3}},$$

when the input is white Gaussian noise with zero mean and unit variance.

Fig. 3 compares the convergence curves of our algorithm with those of other adaptive algorithms in (a) and that of the third weight in (b) with nonwhite Gaussian inputs (SNR = 0 dB). The convergence speeds of the MSE and the third weight for the proposed algorithm are greater than those achieved with that other algorithms and are similar to those in the case where white Gaussian inputs were used. The ANNCFLMS algorithm reaches almost the same convergence speed as that with white Gaussian inputs at high SNR (10 dB; results omitted here). Table 3 lists the experimental misadjustments for nonwhite Gaussian inputs, as is similar to Table 2, the misadjustment of the proposed algorithm is very close to the setting value of 0.01 at the steady-state.

3.3 Correlated Input

Consider the following correlated input [3,5]: both the unknown system (or channel) and the adaptive filter are excited by $x(n) = 0.9x(n-1) + \delta(n)$, where $\delta(n)$ is a zero mean, uncorrelated Gaussian noise of unity variance.

The convergence curves of the MSE with correlated inputs when SNR = 0 dB are shown in Fig. 4. We can observe that the convergence rate of the MSE appears to be the fastest for the proposed method, as is similar to those of with uncorrelated inputs (i.e., white and nonwhite Gaussian inputs presented in Section 3.1 and 3.2). The experimental misadjustments of the algorithm are also listed in Table 4.

3.4 Comparison of Step-size for Various Adaptive Algorithms

In a comparison of the step-size from the perspective of the robustness of low-SNR environments, it can be clearly seen from Table 1 that the experimental results in [2–4,6,8] were sensitive to high-level noise, since the instantaneous error value $e(n)$ was used in their implementations and could, therefore, be contaminated by the noise, while the method in [5] needs the noise signal to be uncorrelated. [8] is attractive compared with [2] and [4] in the experiments of a linear prediction and a nonlinear signal of the speech, but the performance will be degraded on channel estimation problem. Thus, the method used here is no longer reasonable. Wei et al. [7] and Choi et al. [9] exploited the augmented Lagrangian multipliers to enhance the convergence rate. The simulated results show that [7,9] have a better performance.

In the proposed method, it is not only the exploitation of a linear combination (similar to Mathows' (GASS) [4]) which reduces the square estimate error during each iteration, but also the use of a nonlinear update (similar to Mandic's (GNGD) [8]) which deals with the potentially non-Gaussian nature of the data. This enables the step-size to be processed like it is via a combination of both terms. In addition, as can be observed in Fig. 1, the modified cost function always tracks the minimum MSE (i.e., the variance of the estimate noise) and attains its minimum value at steady-state, and therefore, the step-size is also converged into a smaller value to obtain a low misadjustment error. The simulated results demonstrate that the proposed method can perform well and robustly in low-SNR or colored input conditions.

4 Conclusion

This paper introduced a new LMS algorithm, ANNCFLMS. The cost function of the adaptive neural network is used as a criterion of noise estimation to determine how close the adaptive filter attains to optimum performance. The results demonstrate that it is a good approach to unknown channel estimation.

References:

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

[2] A. Benveniste, M. Metivier, P. Priouret, *Adaptive Algorithms and Stochastic Approximation*. New York: Springer-Verlag, 1990.

[3] R. H. Kwong, E. W. Johnston, A variable step size LMS algorithm, *IEEE Trans. Signal Processing*, Vol. 40, July 1992, pp. 1633–1642.

[4] V. J. Mathews, Z. Xie, A stochastic gradient adaptive filter with gradient adaptive step size, *IEEE Trans. Signal Processing*, Vol. 41, June 1993, pp. 2075–2087.

[5] T. Aboulnasr, K. Mayyas, A robust variable step-size LMS-type algorithm: Analysis and simulations, *IEEE Trans. Signal Processing*, Vol. 45, Mar. 1997, pp. 631–639.

[6] W.-P. Ang, B. Farhang-Boroujeny, A new class of gradient adaptive step-size LMS algorithms, *IEEE Trans. Signal Processing*, Vol. 49, No. 4, April 2001, pp. 805–810.

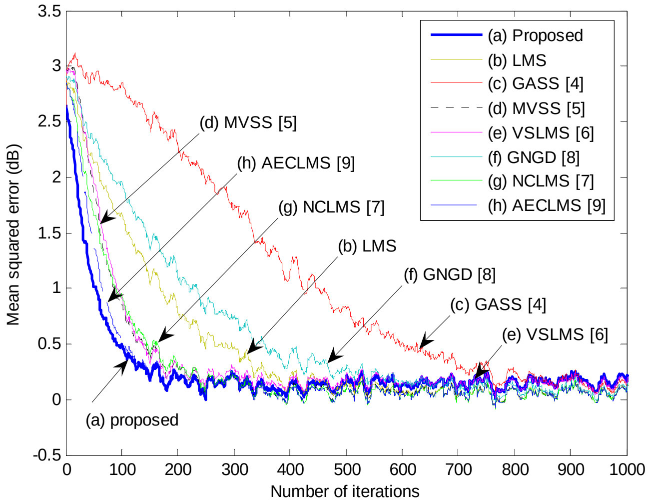
[7] Y. Wei, S. B. Gelfand, J. V. Krogmeier, Noise-constrained least mean squares algorithm, *IEEE Trans. Signal Processing*, Vol. 49, No. 9, Sep. 2001, pp. 1961–1970.

[8] D. P. Mandic, A generalized normalized gradient descent algorithm, *IEEE Signal Process. Lett.*, Vol. 11, No. 2, Feb. 2004, pp. 115–118.

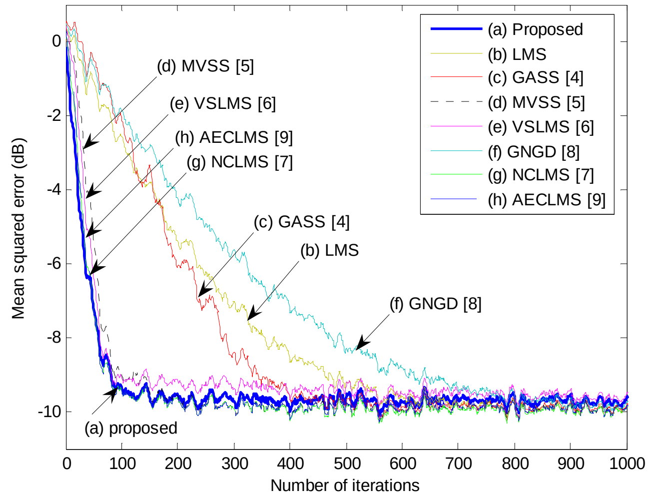
[9] S. Choi, T. Lee, D. Hong, Adaptive error-constrained method for LMS algorithms and applications, *Signal Processing*, Vol. 85, Oct. 2005, pp. 1875–1897.

[10] S. Haykin, *Adaptive Filter Theory*, 4rd ed. Upper Saddle River, NJ: Prentice-Hall, 2002.

[11] D. P. Mandic, J. A. Chambers, *Recurrent Neural Networks for Prediction: Learning Algorithms, Architectures and Stability.* Chichester, U.K.: Wiley, 2001.

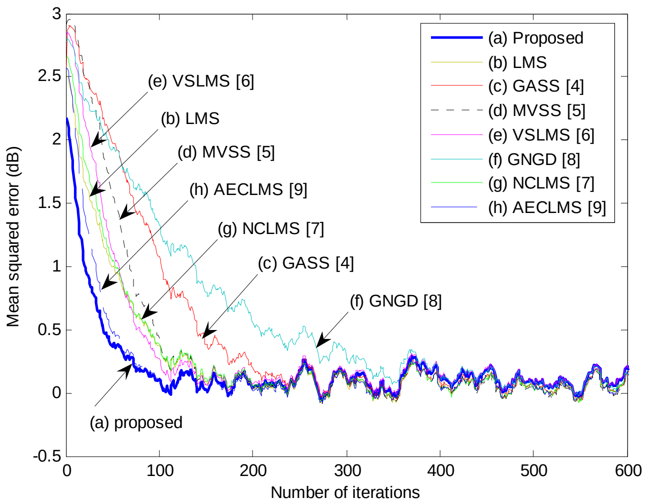


(a)

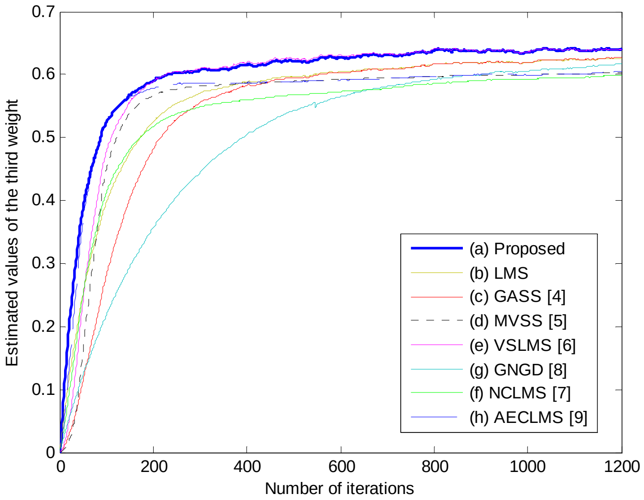


(b)

Fig. 2. Comparison of the convergence curves of the MSE for various adaptive algorithms in the channel estimation problem: (a) SNR = 0 dB and (b) SNR = 10 dB.



(a)



(b)

Fig. 3. Comparison of the convergence curves of (a) the MSE and (b) the third weight for various adaptive algorithms with nonwhite Gaussian inputs when SNR = 0 dB in the channel estimation problem.

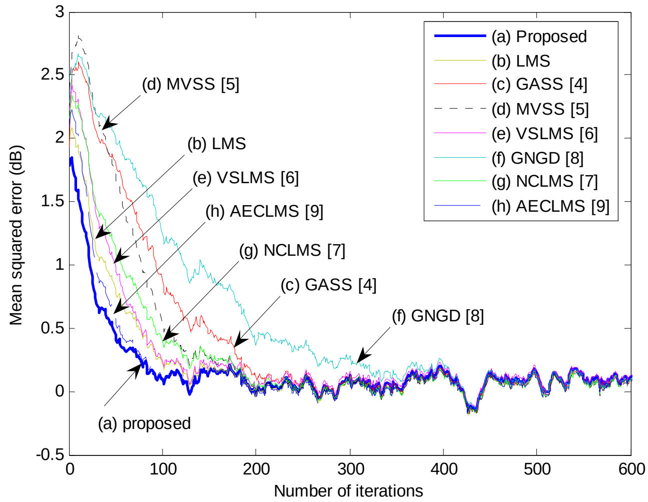


Fig. 4. Comparison of the convergence curves of the MSE for various adaptive algorithms with correlated inputs when SNR = 0 dB in the channel estimation problem.

Table 2. Experiment misadjustment of the adaptive algotrithm for the constant channel with design misadjustment of 0.01 (White Gaussian inputs)

SNR	Misadjustment							
	LMS	GASS	MVSS	VSLMS	NCLMS	GNGD	AECLMS	Proposed
0 dB	0.0094	0.0095	0.0102	0.0112	0.0098	0.0116	0.0098	0.0101
10 dB	0.0098	0.0117	0.0089	0.0122	0.0092	0.0122	0.0094	0.0113

Table 3. Experiment misadjustment of the adaptive algotrithm for the constant channel with design misadjustment of 0.01 (Nonwhite Gaussian inputs)

SNR	Misadjustment							
	LMS	GASS	MVSS	VSLMS	NCLMS	GNGD	AECLMS	Proposed
0 dB	0.0102	0.0112	0.0085	0.0093	0.0082	0.0125	0.0092	0.0099
10 dB	0.0098	0.0142	0.0067	0.0086	0.0062	0.0133	0.0116	0.0101

Table 4. Experiment misadjustment of the adaptive algotrithm for the constant channel with design misadjustment of 0.01 (Correlated inputs)

SNR	Misadjustment							
	LMS	GASS	MVSS	VSLMS	NCLMS	GNGD	AECLMS	Proposed
0 dB	0.0103	0.0115	0.0073	0.0118	0.0113	0.0132	0.0099	0.0107
10 dB	0.0083	0.009	0.0081	0.0122	0.0087	0.0146	0.0092	0.0089