A Least Squares Approach to Escalator Algorithms for Adaptive Filtering

Namyong Kim

In this paper, we introduce an escalator (ESC) algorithm based on the least squares (LS) criterion. The proposed algorithm is relatively insensitive to the eigenvalue spread ratio (ESR) of an input signal and has a faster convergence speed than the conventional ESC algorithms. This algorithm exploits the fast adaptation ability of least squares methods and the orthogonalization property of the ESC structure. From the simulation results, the proposed algorithm shows superior convergence performance.

Keywords: Least squares, escalator, Gram-Schmidt orthogonal, eigenvalue, adaptive algorithm.

Manuscript received Apr. 6, 2005; revised Jan. 20, 2006.

Namyong Kim (phone: + 82 33 570 6400, email: namyong@kangwon.ac.kr) is with Department of Information and Communication Engineering, Kangwon National University, Samcheok, Korea.

I. Introduction

Many researchers have studied various adaptive filter structures and coefficient-adjustment algorithms. The tapped delay line (TDL) filter structure using the least mean square (LMS) algorithm of Widrow and Hoff [1] has been widely used due to its simplicity in realization. One drawback of the LMS algorithm is that its convergence speed decreases as the ratio of the maximum-to-minimum eigenvalues of the input autocorrelation matrix increases. To cope with this problem, it has been proposed to orthogonalize the input signal using the Gram-Schmidt orthogonalization procedure, which can be implemented using the escalator (ESC) structure [2]. Though the ESC structure orthogonalizes the input signals, its coefficient adaptation algorithms have not been sufficiently studied for faster convergence speed. The LMS algorithm for the ESC in which mean squared local estimation errors are minimized is currently used in the ESC filtering problems [3], [4]. In this paper, we present a new fast adaptation algorithm for the ESC structure by introducing the least squares (LS) criterion to the local errors of the structure.

The performance index for the mean squared error (MSE) criterion is defined as the expected value of the squared difference $e(k)$ between the desired information symbol $d(k)$ and the estimated information symbol $y(k)$ at time k:

$$MSE = E[e^2(k)] . \tag{1}$$

In the MSE criterion, the tap weight coefficients of filter structures are adjusted to minimize the MSE. The derivation of the algorithms for adjusting the coefficients of the linear filter to minimize the MSE is based on a statistical approach. Instead of a statistical average, the performance index can be expressed in terms of a time average. We can determine the coefficients of the filter that minimize the time-average weighted square error

$$J(k)=\sum_{n=0}^{k} w^{k-n}e^2(n)\,, \tag{2}$$

where w represents weighting factor $0<w<1$.

By adopting the LS criterion to the ESC structure, minimizing $J(k)$ with respect to the ESC coefficients, we propose an ESC algorithm that has fast convergence speed.

II. Escalator Filter Structure

In a given symmetric matrix, R, there exists a unit lower triangular (ULT) matrix, W, such that WRW^T is a diagonal matrix. The ULT matrix W can be computed in the form of $W=W_N W_{N-1}...W_1$. The ULT transformation $Y(k)=W\cdot X(k)$ means that system W generates the uncorrelated output vector $Y(k)$ for the input vector $X(k)$, where its symmetric autocorrelation matrix $R=E[X(k)X^T(k)]$. If we define $X(k)$ as an input vector augmented with the desired sample $d(k)$, $X(k)=[x(k-N+1),\ x(k-N+2)\ ,...,\ x(k),\ d(k)]^T$ and $Y(k)=[y(k-N+1),\ y(k-N+2)\ ,...,\ y(k),\ e(k)]^T$ as an output vector augmented with the error sample $e(k)$, $Y(k)=W\cdot X(k)$ becomes the filtering process of the ESC structure.

We can realize the ULT transformation stage by stage as $Y_1(k)=W_1\cdot X(k)$, $Y_2(k)=W_2\cdot Y_1(k)$, $Y_3(k)=W_3\cdot Y_2(k)$, and so on. The final stage's output vector $Y_N(k)$ becomes $Y(k)$. For N=3, $X(k)=[x(k-2),x(k-1),x(k),d(k)]^T$, $Y(k)=[y(k-2),y(k-1),y(k),e(k)]^T$, and we can proceed to produce $Y_1(k)$, $Y_2(k)$, and $Y_3(k)$ sequentially as follows:

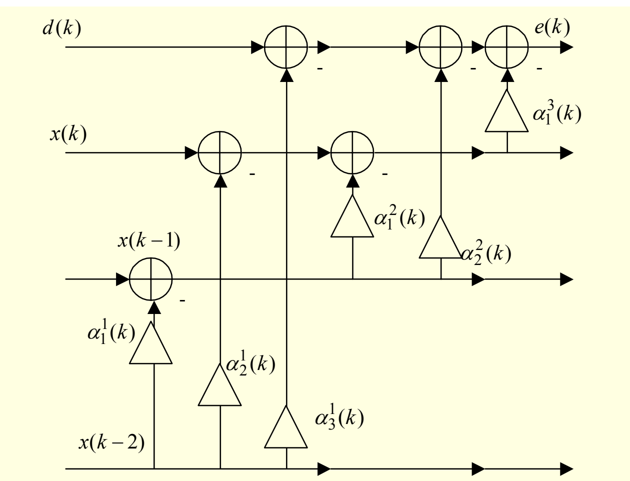


Fig. 1. ESC filter realization for N=3.

$$Y_1(k)=\begin{bmatrix} x(k-2)\\ x(k-1)-\alpha_1^1(k)\cdot x(k-2)\\ x(k)-\alpha_2^1(k)\cdot x(k-2)\\ d(k)-\alpha_3^1(k)\cdot x(k-2)\end{bmatrix}=\begin{bmatrix} y_{10}(k-2)\\ y_{11}(k-1)\\ y_{12}(k)\\ y_{13}(k+1)\end{bmatrix}, \tag{3}$$

$$Y_2(k)=\begin{bmatrix} y_{10}(k-2)\\ y_{11}(k-1)\\ y_{12}(k)-\alpha_1^2(k)\cdot y_{11}(k-1)\\ y_{13}(k+1)-\alpha_2^2(k)\cdot y_{11}(k-1)\end{bmatrix}=\begin{bmatrix} y_{20}(k-2)\\ y_{20}(k-1)\\ y_{21}(k)\\ y_{22}(k+1)\end{bmatrix}, \tag{4}$$

and

$$Y_3(k)=\begin{bmatrix} y_{20}(k-2)\\ y_{20}(k-1)\\ y_{21}(k)\\ y_{22}(k+1)-\alpha_1^3(k)\cdot y_{21}(k)\end{bmatrix}=\begin{bmatrix} y_{30}(k-2)\\ y_{30}(k-1)\\ y_{30}(k)\\ y_{31}(k+1)\end{bmatrix}. \tag{5}$$

The corresponding ESC filter realization for N = 3 is shown in Fig 1. The general ESC filter equations are

$$\begin{aligned} &y_{i,j}(k-m)=y_{i-1,j+1}(k-m)-\alpha_j^i(k)y_{i-1,1}(k-n),\\ &y_{0,N+1}(k+1)=d(k), y_{0,j}(\cdot)=x(\cdot), y_{i,j}(k+1)=e_i(k), \end{aligned} \tag{6}$$

for $1\le i\le n$, $1\le j\le N-i+1$, $m=N-i-j$, and $n=N-i$.

For the escalator structure, the global minimization of the output energy can be accomplished as a sequence of local minimization problems, one at each stage of the escalator filters. Figure 2 shows the part of the escalator filter, which corresponds to the weight $\alpha_j^i(k)$. We can see that part of the escalator filter can be considered as a one-tap coefficient TDL filter.

III. Escalator Coefficient Adaptation by MSE Criterion

The escalator weight $\alpha_j^i(k)$, which can be considered as the coefficient of a one-tap TDL, can be optimized according to the MSE criterion or by employing the method of least squares. Suppose we adopt the MSE criterion and select the parameter to minimize the sum of the mean-square errors where the error is $y_{i,j}(k-m)$. In other words, the MSE is given as (7).

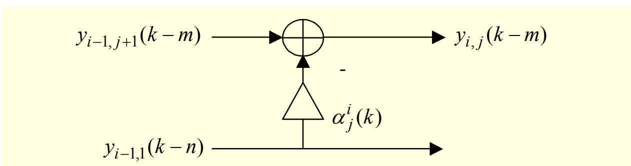


Fig. 2. Part of the escalator filter with $\alpha_j^i(k)$.

$$MSE = E[y^2{}_{i,j}(k-m)] = E[(y_{i-1,j+1}(k-m) - \alpha_j^i(k) y_{i-1,1}(k-n))^2] \quad (7)$$

Differentiation of MSE with respect to $\alpha_j^i(k)$ yields the solution,

$$\alpha_j^i(k) = \frac{E[y_{i-1,j+1}(k-m) y_{i-1,1}(k-n)]}{E[y_{i-1,1}^2(k-n)]}. \quad (8)$$

It is instructive to note that each $\alpha_j^i(k)$ defined is in the form of a ratio whose numerator and denominator are cross-correlation and autocorrelation (variance) terms, respectively.

Thus, the denominator and numerator terms in (8) can be updated as the instantaneous cross-correlation functions via the single-pole low-pass system to yield adaptive escalator filter weights.

$$\alpha_j^i(k) = \frac{c_{i,j}(k)}{v_y{}^i(k)}, \quad (9)$$

where

$$c_{i,j}(k) = \beta c_{i,j}(k-1) + (1-\beta) y_{i-1,j+1}(k-m) y_{i-1,1}(k-n), \quad (10)$$

$$v_y{}^i(k) = \beta v_y{}^i(k-1) + (1-\beta) y_{i-1,1}^2(k-n), \quad (11)$$

and $0 < \beta < 1$. The initial values of $c_{i,j}(k)$ and $v_i{}^i(k)$ are usually zero.

Instead of estimating the cross-correlation functions by using the simple low-pass filter, we can use the steepest descent method with time-varying convergence parameter $\mu_i(k)$ [4] as the following:

$$\alpha_j^i(k+1) = \alpha_j^i(k) + \mu_i(k) y_{i,j}(k-m) y_{i-1,1}(k-n), \quad (12)$$

where $\mu_i(k) = 2\mu / v_y{}^i(k)$ and $v_y{}^i(k)$ is estimated using a recurrence relation given by (11) for $0 < \beta < 1$. Here, the time constant of the LPF is determined by β, whereas the average time constant of the adaptation process of the LMS-type algorithm is dependent on μ. As β becomes smaller, the time constant of the LPF decreases, which yields faster estimation of the signal power but a larger estimation noise in the steady state. For the sake of convenience, we call algorithm (12) LMS-ESC in this paper.

IV. Escalator Coefficient Adaptation by LS Criterion

The use of the least squares method leads to developing a recursive algorithm for the design of adaptive filters. The resulting algorithm is referred to as the recursive least squares (RLS) algorithm [5]. An important feature of the RLS algorithm is that it provides significantly faster convergence than the LMS algorithm.

In this section, we propose to adopt the LS criterion to the local ESC filter structure for updating $\alpha_j^i(k)$. From (2) and Fig. 2, the performance index to be minimized is

$$J(k) = \sum_{p=0}^{k} w^{k-p} y_{ij}^2(p-m) = \sum_{p=0}^{k} w^{k-p} [y_{i-1,j+1}(p-m) - y_{i-1,1}(p-n) \cdot \alpha_j^i(k)]^2. \quad (13)$$

Minimization of $J(k)$ with respect to $\alpha_j^i(k)$ yields the solution

$$\alpha_j^i(k) = \frac{\sum_{p=0}^{k} w^{k-p} y_{i-1,j+1}(p-m) y_{i-1,1}(p-n)}{\sum_{p=0}^{k} w^{k-p} y_{i-1,1}^2(p-n)}. \quad (14)$$

The numerator and denominator in (14) can be updated recursively in time as follows:

$$A(k) = w \cdot A(k-1) + y_{i-1,j+1}(k-m) y_{i-1,1}(k-n) \quad (15)$$

$$B(k) = w \cdot B(k-1) + y_{i-1,1}^2(k-n). \quad (16)$$

Then,

$$\alpha_j^i(k) = A(k) B^{-1}(k). \quad (17)$$

The inverse of $B(k)$ can be rearranged as follows:

$$\begin{aligned} B^{-1}(k) &= [w \cdot B(k-1) + y_{i-1,1}^2(k-n)]^{-1} \\ &= w^{-1} \cdot B^{-1}(k-1) \left[\frac{w}{w + y_{i-1,1}^2(k-n) B^{-1}(k-1)}\right] \\ &= w^{-1} \cdot \left[B^{-1}(k-1) - \frac{y_{i-1,1}^2(k-n) B^{-2}(k-1)}{w + y_{i-1,1}^2(k-n) B^{-1}(k-1)}\right]. \end{aligned} \quad (18)$$

Thus, $B^{-1}(k)$ can be computed recursively according to (18). In (18), it is convenient to define $g(k)$ as

$$g(k) = \frac{y_{i-1,1}(k-n) B^{-1}(k-1)}{w + y_{i-1,1}^2(k-n) B^{-1}(k-1)}. \quad (19)$$

Then, (18) becomes

$$B^{-1}(k) = w^{-1} \cdot [B^{-1}(k-1) - g(k) y_{i-1,1}(k-n) B^{-1}(k-1)]. \quad (20)$$

By multiplying both sides of (20) by $y_{i-1,1}(k-n)$, we can acquire

$$B^{-1}(k)y_{i-1,1}(k-n)=w^{-1}[B^{-1}(k-1)y_{i-1,1}(k-n) - g(k)y_{i-1,1}^2(k-n)B^{-1}(k-1)]=g(k). \tag{21}$$

Substituting (15) into (17) and using (20) yields

$$\begin{aligned}\alpha_j^i(k) &= A(k)B^{-1}(k)\\ &= w^{-1}\cdot[B^{-1}(k-1)-g(k)y_{i-1,1}(k-n)B^{-1}(k-1)]\\ &\quad \cdot[w\cdot A(k-1)+y_{i-1,j+1}(k-m)y_{i-1,1}(k-n)]\\ &= B^{-1}(k-1)A(k-1)\\ &\quad -g(k)y_{i-1,1}(k-n)B^{-1}(k-1)A(k-1)\\ &\quad +w^{-1}B^{-1}(k-1)y_{i-1,j+1}(k-m)y_{i-1,1}(k-n)\\ &\quad -w^{-1}g(k)y_{i-1,j+1}(k-m)y_{i-1,1}^2(k-n)B^{-1}(k-1)\\ &= w^{-1}[B^{-1}(k-1)-g(k)y_{i-1,1}(k-n)B^{-1}(k-1)]\\ &\quad \cdot[wA(k-1)+y_{i-1,j+1}(k-m)y_{i-1,1}(k-n).\end{aligned} \tag{22}$$

Rearranging (22) into a time-recursive equation of $\alpha_j^i(k)$, (22) produces

$$\begin{aligned}&\alpha_j^i(k)\\ &=\alpha_j^i(k-1)-g(k)y_{i-1,1}(k-n)\alpha_j^i(k-1)\\ &+y_{i-1,j+1}(k-m)y_{i-1,1}(k-n)\\ &\cdot w^{-1}[B^{-1}(k-1)-B^{-1}(k-1)g(k)y_{i-1,1}(k-n)].\end{aligned} \tag{23}$$

Substituting (20) into (23) and using (21) make (23) into the following:

$$\begin{aligned}\alpha_j^i(k)&=\alpha_j^i(k-1)-g(k)y_{i-1,1}(k-n)\alpha_j^i(k-1)+y_{i-1,j+1}(k-m)g(k)\\ &=\alpha_j^i(k-1)+g(k)[y_{i-1,j+1}(k-m)-y_{i-1,1}(k-n)\alpha_j^i(k-1)].\end{aligned} \tag{24}$$

Noting that $y_{i-1,j+1}(k-m)-y_{i-1,1}(k-n)\alpha_j^i(k-1)$ is the error, $y_{i,j}(k-m)$, $\alpha_j^i(k)$ is updated recursively according to the relation (which we call RLS-ESC in this paper),

$$\alpha_j^i(k)=\alpha_j^i(k-1)+g(k)y_{i,j}(k-m), \tag{25}$$

Table 1. Comparison of the number of multiplications per output sample.

Algorithm	Number of multiplications and divisions
LMS-TDL	$2N+1$
RLS-TDL	$2.5N^2+4.5N+2$
LMS-ESC	$3(N^2+N)$
RLS-ESC	$4(N^2+N)$

Note: N is the number of taps in TDL and the number of stages in ESC.

where $g(k)$ and $B^{-1}(k)$ are computed by (19) and (20), respectively. The initial value of $B(k)$ is a small positive constant to avoid $B(k)$ from being ill-conditioned.

The performance improvement of the RLS algorithm is achieved at the expense of a large increase in computational complexity. For the RLS algorithm in the transversal filter (RLS-TDL), the number of computations (multiplications and divisions) is proportional to N^2 [6]. The ESC filter structure with N stages has $0.5(N^2+N)$ coefficients. From (11) and (12), LMS-ESC requires 5 multiplications and 1 division for the update of the local coefficient $\alpha_j^i(k)$. So, the number of computations per output sample in LMS-ESC with N stages is $3(N^2+N)$. The number of computations in RLS-ESC is $4(N^2+N)$, which can be computed in (19), (20), and (25). Notice that the adaptive algorithms for ESC do not have a large difference in complexity, but the ESC structure itself has a drawback of large computational complexity, which is proportional to N^2. Table 1 lists the computational complexity of these adaptive algorithms. The corresponding graphs of the number of computations as a function of filter stages (taps in TDL) are shown in Fig. 3.

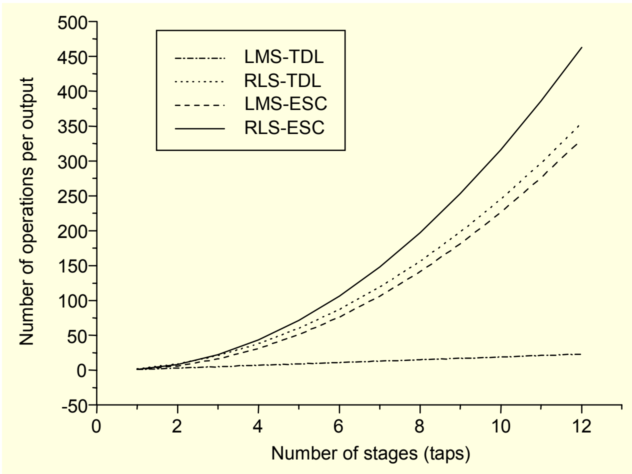


Fig. 3. Computational complexity of adaptive algorithms.

V. Results and Discussions

1. Channel Equalization

In this section, the performances of adaptive equalization for two channels that have different eigenvalue spread ratio (ESR) values, and the performances of system identification for four different inputs, are compared. The algorithms considered are the LMS-TDL algorithm [1], LMS-ESC algorithm of (12), and RLS-ESC in (25). The equalization performance is investigated in polar form with $d(k)=\pm 1$. All values are equiprobable. The random input signal $d(k)$ is transmitted to the channel, and the channel output is corrupted by the white Gaussian noise sequence. The adaptive equalizer has the task of correcting for the distortion produced by the channel in the presence of additive white noise. The impulse response h_i of the channel model is

$$h_i = \frac{1}{2}\{1+\cos[2\pi(i-2)/BW]\}, \quad i = 1,2,3, \qquad (26)$$

where the parameter BW determines the channel bandwidth and controls the eigenvalue spread ratio of the correlation matrix of the inputs in the equalizer [5]. The eigenvalue spread ratio increases with BW. The experiment is carried out in two channels that are intended to evaluate the convergence performance of the algorithms using the TDL and ESC structures to changes in the eigenvalue spread.

Channel 1: BW=3.1, ESR=11.12, and
Channel 2: BW=3.3, ESR=21.71.

The LMS-TDL equalizer has 11 tap coefficients. The LMS-ESC and the proposed algorithm consist of 11 stages. A zero mean white Gaussian noise sequence with variance 0.001 was added to yield the equalizer input. The convergence parameter for the LMS-TDL was 2μ=0.02. The convergence parameter 2μ and smoothing parameter β for LMS-ESC are 0.02 and 0.99, respectively. The weighting factor $\boldsymbol{w}$ for the proposed algorithms based on the LS criterion is 0.99. These convergence parameter values for those algorithms were determined experimentally so that the related steady-state mean squared errors were the same. MSE learning curves were obtained by ensemble averaging over 500 independent trials of the experiment.

We see in Figs. 4 and 5 for equalization that increasing the ESR has the effect of increasing the steady-state MSE value of the algorithms and slowing down the rate of the convergence of the LMS-TDL equalizer algorithm. In channel 1, approximately 900 samples are required for the LMS-TDL to converge, and in channel 2, LMS-TDL requires far more than 900 samples to converge. On the other hand, the algorithms having the ESC structure show no slowing down of their convergence speed. In both cases, the proposed algorithm has shown more rapid convergence than the LMS-ESC. The LMS-ESC converges after about 300 samples, but RLS-ESC requires about 100 samples. The steady-state values of ESC algorithms are relatively insensitive to variations in ESR.

In both channels, LMS-TDL is subject to severe performance degradation and has shown much slower convergence in channel 2. The ESC algorithms show no slowing in their convergence speed in both channels. Though increasing the ESR has the effect of increasing the steady-state MSE value of the algorithms, the steady-state values of ESC algorithms are relatively insensitive to ESR variations

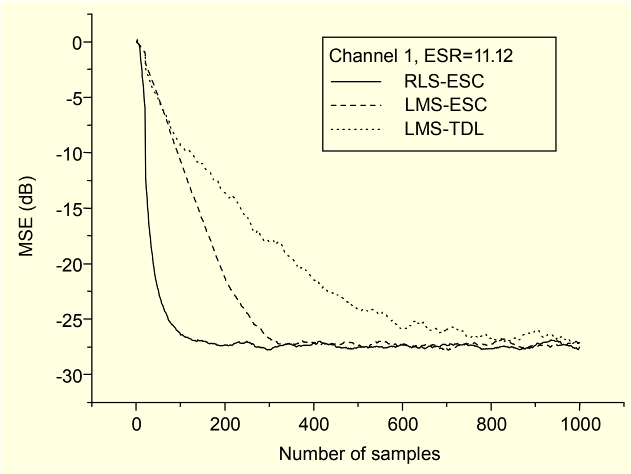


Fig. 4. Convergence performance of equalization for eigenvalue spread ratio (ESR) = 11.12.

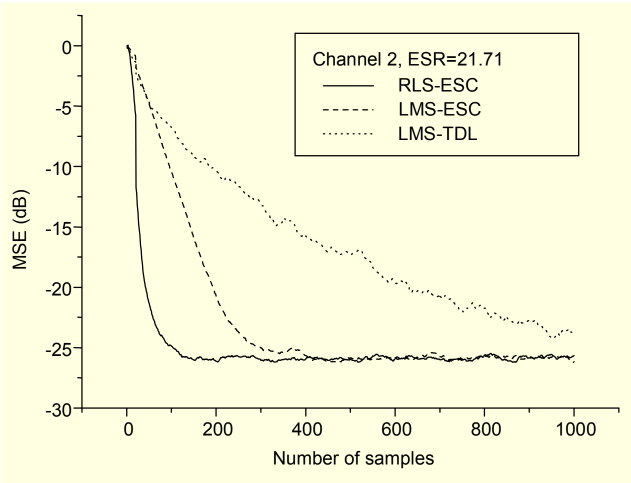


Fig. 5. Convergence performance of equalization for eigenvalue spread ratio (ESR) = 21.71.

compared to the LMS-TDL.

2. System Identification

In system identification applications of adaptive filtering, the desired signal is derived by passing the input through a finite impulse response (FIR) filter (the unknown system) of length 9. The impulse response of the unknown system is chosen to follow a triangular wave form that is symmetric with respect to the center tap point [5]. The order of the adaptive filters is also N=9. The additive white noise, uncorrelated with the input, is added to the output of the unknown system so that the signal-to-noise ratio is 100 dB. The inputs are generated by filtering white noise with one of four order-32 linear phase FIR noise-coloring filters. The frequency magnitude responses of the coloring filters are shown in Figs. 6(a) through 6(d) [4]. These

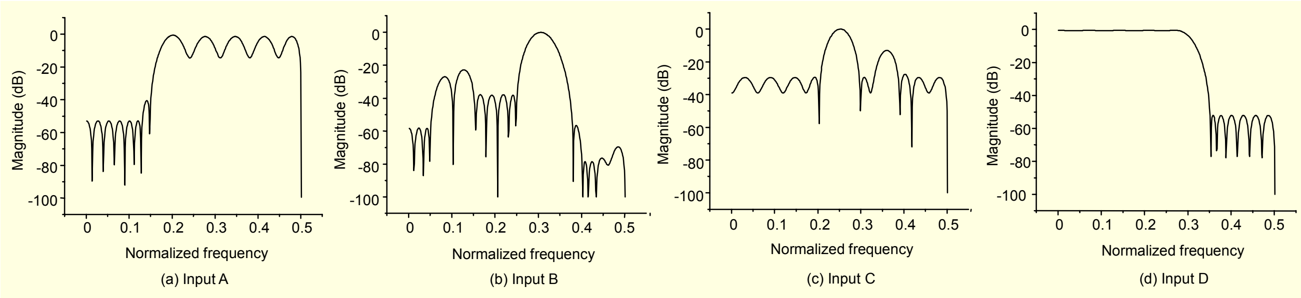


Fig. 6. Frequency magnitude responses of the coloring filters.

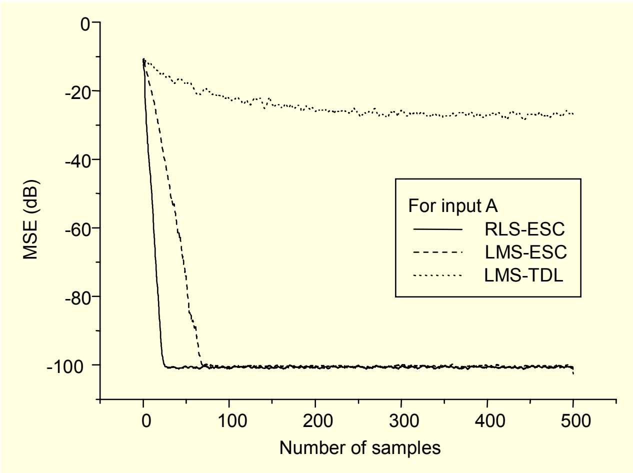


Fig. 7. MSE learning curves for the system identification simulations with input A.

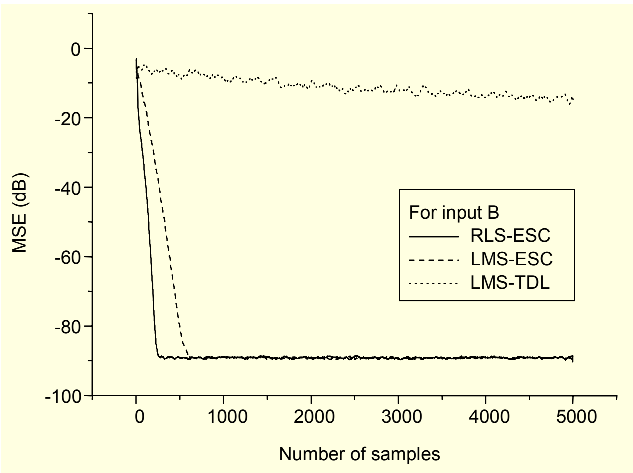


Fig. 8. MSE learning curves for the system identification simulations with input B.

coloring filters generate considerable variations in the input statistics required to verify the performance of the proposed algorithm. MSE learning curves were obtained by ensemble averaging over 500 independent trials of the experiment, as in the simulation of channel equalization.

The MSE learning curves for the four inputs are shown in Figs. 7 through 10. The convergence parameter for the LMS-TDL was 2μ=0.002. The convergence parameter 2μ and smoothing parameter β for LMS-ESC are 0.02 and 0.98, respectively. The weighting factor w for RLS-ESC is 0.98.

For all the input signals, the proposed RLS-ESC converges faster than the LMS-ESC. The LMS-TDL shows that it can not reach below -40 dB for any input, whereas the algorithms having the ESC structure converge rapidly to around -100 dB. The superior performance of the proposed algorithm can also be expected in our further research including echo cancellation problems because the echo cancellation with speech signals can be modeled as an adaptive system identification problem.

In Fig. 11, MSE learning curves of RLS-ESC are shown to evaluate steady state values for the four different inputs. For the high-pass and low-pass input signals, the RLS-ESC converges to -100 dB, but for the bandpass input signals, it shows a small variation of steady state value up to 15 dB. These results indicate that the ESC structure gives fast convergence but still requires some methods to consolidate its performance robustness under varying input conditions.

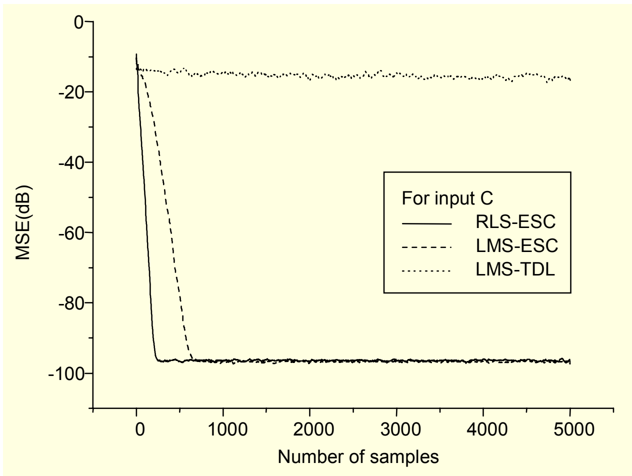


Fig. 9. MSE learning curves for the system identification simulations with input C.

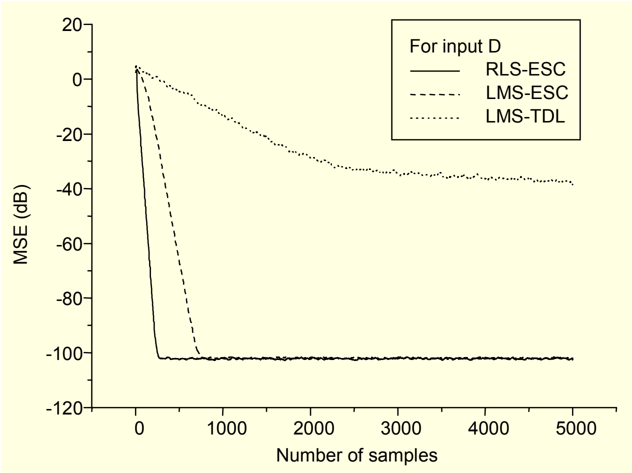


Fig. 10. MSE learning curves for the system identification simulations with input D.

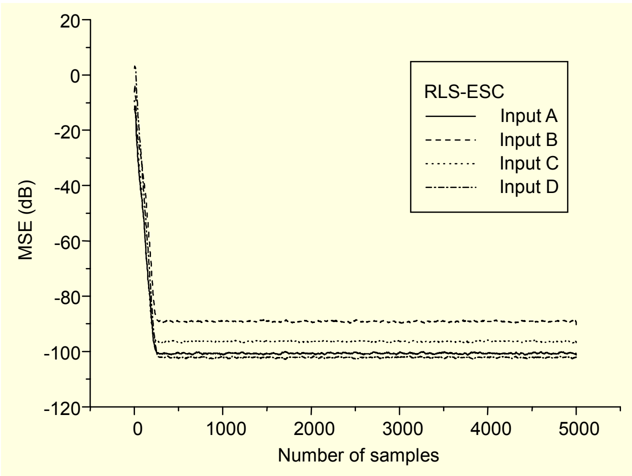


Fig. 11. MSE learning curves for RLS-ESC with four inputs.

VI. Conclusions

In this paper, the anthor proposed a new escalator-coefficient adaptation algorithm that is relatively independent of the eigenvalue spread ratio and has faster convergence speeds than the conventional ESC algorithms. This algorithm adopts the least squares criterion to the ESC structure. It exploits the fast adaptation ability of the LS method and the orthogonalization property of the ESC structure. The simulation results show that the proposed algorithm has faster convergence speed than the LMS-ESC algorithm and it shows no slowing down of convergence speed when we increase the eigenvalue spread ratio of the channel. The steady-state values of ESC algorithms are relatively insensitive to ESR variations compared to the LMS-TDL. The improved performance indicates that the proposed algorithm does appear to be an attractive alternative to the MSE-criterion ESC algorithms for adaptive filtering with large eigenvalue disparity. But in system identification applications with the bandpass input signals, the proposed algorithm shows a small variation of steady state value. This indicates that the ESC structure still needs further research for performance robustness under varying input conditions.

Acknowledgment

The author would like to thank the unknown referees for their helpful comments and suggestions about this work.

References

[1] B. Widrow, J.M. McCool, M.G. Larimore, and C.R. Johnson, "Stationary and Nonstationary Learning Characteristics of the LMS Adaptive Filter," *Proc. IEEE*, vol. 64, Aug. 1976, pp. 1151-1162.

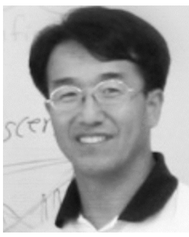
[2] K.N. Ahmed and D.H. Youn, "On Realization and Related Algorithm for Adaptive Prediction," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-28, Oct. 1980, pp. 493-497.

[3] K.M. Kim, I.W. Cha, and D.H. Youn, "Adaptive Multichannel Digital Filter with Lattice-Escalator Hybrid Structure," *Proc. of the 1990 ICASSP*, New Mexico, USA, vol. 3, Apr. 1990, pp. 1413-1416.

[4] V.N. Parikh and A.Z. Baraniecki, "The Use of the Modified Escalator Algorithm to Improve the Performance of Transform-Domain LMS Adaptive Filters." *IEEE Trans. on Signal Processing*, vol. 46, no. 3, Mar. 1998, pp. 625-635.

[5] S. Haykin, *Adaptive Filter Theory*, 2nd ed., Englewood Cliffs, NJ: Prentice Hall, 1991.

[6] J.G. Proakis, *Digital Communications*, 2nd ed., McGraw-Hill, 1989,



Namyong Kim received the BS, MS, and PhD degrees from Yonsei University in 1986, 1988, and 1991, all in electronic engineering. From 1992 to 1997, he was with Kwandong University in Kangnung. Since 1998, he has been an Associate Professor in the Department of Information & Communications Engineering, Samcheok National University. His current research interests are in adaptive signal processing in mobile communications and in odor sensing-identification systems.