Signal Processing 26 (1992) 335–358
Elsevier

Constrained adaptive LMS L-filters*

C. Kotropoulos and I. Pitas
Department of Electrical Engineering, University of Thessaloniki, Thessaloniki 54006, Greece

Received 14 March 1991
Revised 26 July 1991 and 24 October 1991

Abstract. Two novel adaptive nonlinear filter structures are proposed which are based on linear combinations of order statistics. These adaptive schemes are modifications of the standard LMS algorithm and have the ability to incorporate constraints imposed on coefficients in order to permit location-invariant and unbiased estimation of a constant signal in the presence of additive white noise. The convergence in the mean and in the mean square of the proposed adaptive nonlinear filters is studied. The rate of convergence is also considered. It is verified by simulations that the independence theory provides useful bounds on the rate of convergence. The extreme eigenvalues of the matrix which controls the performance of the location-invariant adaptive LMS L-filter are related to the extreme eigenvalues of the correlation matrix of the ordered noise samples which controls the performance of other adaptive LMS L-filters proposed elsewhere. The proposed filters can adapt well to a variety of noise probability distributions ranging from the short-tailed ones (e.g. uniform distribution) to long-tailed ones (e.g. Laplacian distribution).

Zusammenfassung. Es werden zwei neue adaptive nichtlineare Filterstrukturen vorgeschlagen, die auf Linearkombinationen von Order-Statistik beruhen. Diese adaptiven Strukturen sind Modifikationen des üblichen LMS-Algorithmus und erlauben die Einbringung von Bedingungen bezüglich der Koeffizienten, um ortsinvariante und erwartungstreue Schätzungen konstanter Signale unter additivem, weißem Rauschen zu ermöglichen. Die Konvergenz bezüglich des Mittelwertes und des quadratischen Mittelwertes wird für die vorgeschlagenen nichtlinearen Filter untersucht. Weiterhin wird die Konvergenzgeschwindigkeit betrachtet. Durch Simulationen wird gezeigt, daß die Independence-Theorie brauchbare Grenzen für die Konvergenzrate liefert. Die extremen Eigenwerte der Matrix, die das Verhalten des ortsinvarianten adaptiven LMS L-Filters bestimmt, werden den Eigenwerten der Korrelationsmatrix der geordneten Rauschabtastwerte gegenübergestellt, die das Verhalten anderer adaptiver LMS L-Filter bestimmt. Die vorgeschlagenen Filter stellen sich sehr gut auf eine Vielzahl verschiedener Verteilungsdichten des Rauschens ein, angefangen von schmalen Verteilungen (z.B. Gleichverteilung) bis hin zu langsam abfallenden (z.B. Laplace).

Résumé. Nous proposons deux structures de filtre non-linéaire originales, structures basées sur des combinaisons linéaires de statistiques d'ordre. Ces techniques adaptatives sont des modifications de l'algorithme LMS standard et ont la capacité d'incorporer des contraintes imposées sur les coefficients afin de permettre une estimation ne variant pas selon la localisation et non biaisée d'un signal constant en présence de bruit blanc additif. Nous étudions la convergence en moyenne et en moyenne quadratique des filtres non-linéaires adaptatifs proposés. Nous considérons également le taux de convergence. Nous vérifions par des simulations que l'hypothèse d'indépendance fournit des bornes utiles sur le taux de convergence. Nous relions les valeurs propres extrêmes de la matrice qui contrôle les performances du L-filtre LMS adaptatif ne variant pas selon la localisation aux valeurs propres extrêmes de la matrice de corrélation des échantillons de bruit ordonnés qui contrôle les performances d'autres L-filtres LMS proposés ailleurs. Les filtres proposés peuvent s'adapter aisément à une variété de distributions de densité de bruit allant de celles à queue courte (p.e. la distribution uniforme) à celles à queue longue (p.e. la distribution de Laplace).

Keywords. Adaptive filters, nonlinear filters, order statistics, L-filters, LMS adaptation.

1. Introduction

Adaptive filters constitute an important part of statistical signal processing. They offer an attractive solution whenever there is a requirement to process signals that result from operation in an environment

* This work was supported by the National Scholarship Foundation of Greece and by the Bodosaki Foundation.

of unknown statistics and they provide a significant improvement in performance over the use of non-adaptive filters designed by conventional methods (e.g. Wiener filter). They have been applied in a wide variety of problems, for example in system identification, adaptive equalization for data transmission, digital representation of speech, adaptive autoregressive spectrum analysis, adaptive detection of a signal in noise of unknown statistics, echo cancellation [2, 11].

Many algorithms for nonlinear filters have appeared in the literature for one-dimensional and two-dimensional signal filtering [22]: Volterra filters, memoryless nonlinearities followed by a linear combiner, order statistic filters (e.g. L-filters), filters that combine in a unique class the L-filters and the linear filters (Ll-filters, median hybrid filters). Each estimator provides a reasonable solution only to specific filtering problems and often the choice of the filter is determined by experience.

Numerous researchers have attempted to combine adaptive filtering and nonlinear filtering. Nonlinear adaptive Volterra filters have been proposed in the cases where the transmission channel is nonlinear for echo cancellation [1, 26]. Extensions of the LMS and RLS algorithms to the adaptation of L-filter and Ll-filter coefficients have been proposed in [18, 21, 23]. The backpropagation algorithm has been used in the adaptation of the coefficients of the hybrid linear order statistic filters, median hybrid filters and Ll-filters in [16]. An adaptive computation of the coefficients of the recursive Ll-filters has been proposed in [17]. The performance of the median LMS, average LMS and trimmed-mean LMS filters is analyzed in [10]. The use of the LMS algorithm for the adaptation of the coefficients of the weighted median is proposed in [24].

The main purpose of this paper is to extend the standard LMS algorithm by applying it to the adaptation of the L-filter coefficients in order to incorporate constraints imposed on the coefficients. L-filters are defined as linear combinations of the ordered data in the filter window. In [5, 19] it is proven that the optimal L-filter for the estimation of a constant signal corrupted by zero mean additive white noise should be either location-invariant or unbiased. Thus, it is reasonable to incorporate the constraints underlying the location-invariant and unbiased estimation to the LMS adaptation algorithm. The use of an adaptive formula for the L-filter coefficients eliminates the need to compute the correlation matrix of the ordered noise samples. It also leads to a filter which easily adapts to an unknown noise probability density function. The LMS algorithm is chosen because of its simplicity and its low computational complexity. It will be seen later on that the incorporation of the above-mentioned constraints to the adaptation procedure eliminates the component with the slowest convergence rate and accelerates the convergence of the L-filter coefficients to the optimal ones. The design of an L-filter which minimizes the mean square error between some desired response and the actual filter output is reviewed. The case of a constant signal corrupted by zero mean additive white noise is considered. The constraints underlying the location-invariant and unbiased estimation are stated. Two novel adaptive schemes are derived by rewriting the normal equations in a form that takes into account the above-mentioned constraints and by using instantaneous values for the correlations of the ordered noise samples. The convergence in the mean of these algorithms to the optimal filter is proven. The convergence in the mean square is also studied. The extreme eigenvalues of the matrix controlling the adaptation in the location-invariant adaptive LMS L-filter are related to the extreme eigenvalues of the correlation matrix of the ordered noise samples. It is verified by simulations that the slowest rate of convergence of the location-invariant adaptive LMS L-filter is smaller than the corresponding one of the adaptive LMS L-filters proposed in [18, 21,23] which are based on the correlation matrix of the ordered noise samples. It is proven that the proposed adaptive filter structures can adapt well in a variety of noise probability distributions ranging from short-tailed ones (e.g. uniform distribution) to long-tailed ones (e.g. Laplacian distribution).

The outline of this paper is as follows. The description of L-filters and the location-invariant and unbiased estimation are given in Section 2. The mathematical derivation of the proposed adaptive L-filters is analyzed in Section 3. Their convergence properties and theoretical bounds on the convergence rate are described in Section 4. Simulation examples of the performance of the proposed adaptive nonlinear filters are included in Section 5. Conclusions are drawn in Section 6.

2. Optimal L-filter for a constant signal in white noise

A constant signal s corrupted by zero-mean additive white noise is considered. The input samples are given by

$$x_i = s + n_i, \tag{1}$$

where n_i are independent, identically distributed random variables satisfying $\mathrm{E}[n_i]=0$. We shall assume that the noise distribution is symmetric about zero. Let $\boldsymbol{x}(k)$ denote the $M\times 1$ input vector at time instant k, i.e.,

$$\boldsymbol{x}(k) = [x_1^k, x_2^k, \ldots, x_M^k]^{\mathrm{T}} = [x_{k-(M-1)/2}, \ldots, x_k, \ldots, x_{k+(M-1)/2}]^{\mathrm{T}}, \tag{2}$$

where M is assumed to be odd. If the data are arranged in ascending order according to their magnitude, the order statistics result:

$$x_{(1)}^k \leqslant x_{(2)}^k \leqslant \cdots \leqslant x_{(M)}^k, \tag{3}$$

where $x_{(i)}^k$ is the i-th largest observation data (i-th order statistic). Thus, the ordered tap-input vector at time instant k is

$$\boldsymbol{x}_r(k) = [x_{(1)}^k, x_{(2)}^k, \ldots, x_{(M)}^k]^{\mathrm{T}}. \tag{4}$$

The output of the L-filter at time k is given by

$$y(k) = \boldsymbol{a}^{\mathrm{T}} \boldsymbol{x}_r(k), \tag{5}$$

where $\boldsymbol{a} = [a_1, \ldots, a_M]^{\mathrm{T}}$ denotes the vector of the L-filter coefficients.

Two methods for estimating the constant signal s from the ordered tap-input vector have been proposed in the literature [5, 19]:
(a) location-invariant estimation,
(b) unbiased estimation.
In the following the conditions for each type of estimation are summarized because they will be used in the derivation of the adaptation formulae. Let $\boldsymbol{e}_M$ denote the $M\times 1$ unitary vector, i.e. $\boldsymbol{e}_M = [1, 1, \ldots, 1]^{\mathrm{T}}$.

LEMMA 1. The necessary and sufficient condition for a location-invariant L-filter which is used in the estimation of a constant signal in white noise is

$$\boldsymbol{e}_M^{\mathrm{T}} \boldsymbol{a} = 1. \tag{6}$$

LEMMA 2. The sufficient conditions for an unbiased L-filter which is used in the estimation of a constant signal in white noise are

$$\boldsymbol{e}_M^{\mathrm{T}} \boldsymbol{a} = 1, \qquad a_j = a_{M-j+1}, \quad j = 1, \ldots, (M-1)/2. \tag{7}$$

The noise distribution should be symmetric about zero.

The coefficients of the L-filter are chosen so that the mean square error (MSE)

$$J=\mathrm{E}[(y(k)-s)^2] \tag{8}$$

is minimized. The location-invariant/unbiased L-filter $\boldsymbol{a}$ is given by [5, 19]

$$\boldsymbol{a}=\frac{\boldsymbol{R}^{-1}\boldsymbol{e}_M}{\boldsymbol{e}_M^{\mathrm{T}}\boldsymbol{R}^{-1}\boldsymbol{e}_M}, \tag{9}$$

where $\boldsymbol{R}$ is the correlation matrix of the ordered noise samples with (i,j) element $r_{ij}=\mathrm{E}[n_{(i)}n_{(j)}]$, $i,j=1,\ldots,M$. An order-recursive algorithm for the calculation of the elements of $\boldsymbol{R}$ can be derived by exploiting recurrence relations on the moments of the order statistics [4, 6, 9, 25, 27].

3. Constrained LMS adaptive L-filters

3.1. *Location-invariant LMS adaptive* L-filter

Let $\boldsymbol{n}_r$ denote the vector of the ordered noise samples, i.e.,

$$\boldsymbol{n}_r=[n_{(1)},n_{(2)},\ldots,n_{(M)}]^{\mathrm{T}}, \tag{10}$$

where the time index k is dropped out for notation simplicity. The MSE J, given in (8), is rewritten as

$$J=\mathrm{E}[\boldsymbol{a}^{\mathrm{T}}\boldsymbol{n}_r\boldsymbol{n}_r^{\mathrm{T}}\boldsymbol{a}]=\boldsymbol{a}^{\mathrm{T}}\mathbf{R}\boldsymbol{a}. \tag{11}$$

Let $\boldsymbol{e}$ denote the $(M-1)/2\times 1$ unitary vector $\boldsymbol{e}_{(M-1)/2}$. The coefficient vector $\boldsymbol{a}$ can be rewritten in the following form by using (6):

$$\boldsymbol{a}=[\boldsymbol{a}_1^{\mathrm{T}}\,|\,1-\boldsymbol{e}^{\mathrm{T}}\boldsymbol{a}_1-\boldsymbol{e}^{\mathrm{T}}\boldsymbol{a}_2\,|\,\boldsymbol{a}_2^{\mathrm{T}}]^{\mathrm{T}}, \tag{12}$$

where $\boldsymbol{a}_1$, $\boldsymbol{a}_2$ are $(M-1)/2\times 1$ vectors given by

$$\boldsymbol{a}_1=[a_1,\ldots,a_{(M-1)/2}]^{\mathrm{T}},\quad \boldsymbol{a}_2=[a_{(M+3)/2},\ldots,a_M]^{\mathrm{T}}. \tag{13}$$

Similarly, the vector of the ordered noise samples can also be partitioned:

$$\boldsymbol{n}_r=[\boldsymbol{n}_{r1}^{\mathrm{T}}\,|\,n_{((M+1)/2)}\,|\,\boldsymbol{n}_{r2}^{\mathrm{T}}]^{\mathrm{T}}, \tag{14}$$

where

$$\boldsymbol{n}_{r1}=[n_{(1)},\ldots,n_{((M-1)/2)}]^{\mathrm{T}},\qquad \boldsymbol{n}_{r2}=[n_{((M+3)/2)},\ldots,n_{(M)}]^{\mathrm{T}}. \tag{15}$$

The correlation matrix $\boldsymbol{R}$ is partitioned as follows:

$$\boldsymbol{R}=\begin{bmatrix}\boldsymbol{R}_1 & \boldsymbol{r}_1 & \boldsymbol{R}_2\\ \boldsymbol{r}_1^{\mathrm{T}} & r & \boldsymbol{r}_2^{\mathrm{T}}\\ \boldsymbol{R}_3 & \boldsymbol{r}_2 & \boldsymbol{R}_4\end{bmatrix}, \tag{16}$$

where

$$\begin{gathered}\boldsymbol{R}_1=\mathrm{E}[\boldsymbol{n}_{r1}\boldsymbol{n}_{r1}^{\mathrm{T}}],\quad \boldsymbol{R}_2=\mathrm{E}[\boldsymbol{n}_{r1}\boldsymbol{n}_{r2}^{\mathrm{T}}],\quad \boldsymbol{R}_3=\mathrm{E}[\boldsymbol{n}_{r2}\boldsymbol{n}_{r1}^{\mathrm{T}}],\quad \boldsymbol{R}_4=\mathrm{E}[\boldsymbol{n}_{r2}\boldsymbol{n}_{r2}^{\mathrm{T}}],\\ \boldsymbol{r}_1=\mathrm{E}[n_{((M+1)/2)}\boldsymbol{n}_{r1}],\quad \boldsymbol{r}_2=\mathrm{E}[n_{((M+1)/2)}\boldsymbol{n}_{r2}]\quad\text{and}\quad r=\mathrm{E}[n_{((M+1)/2)}^2].\end{gathered} \tag{17}$$

It is recognized that

$$R_1 = R_1^{\mathrm{T}}, \quad R_4 = R_4^{\mathrm{T}}, \quad R_3 = R_2^{\mathrm{T}}. \tag{18}$$

The MSE (11) is rewritten as

$$J = r - 2\tilde{a}^{\mathrm{T}} p' + \tilde{a}^{\mathrm{T}} R' \tilde{a}, \tag{19}$$

where

$$\tilde{a} = [a_1^{\mathrm{T}} \mid a_2^{\mathrm{T}}]^{\mathrm{T}}, \quad p' = [re^{\mathrm{T}} - r_1^{\mathrm{T}} \mid re^{\mathrm{T}} - r_2^{\mathrm{T}}]^{\mathrm{T}}, \tag{20}$$

$$R' = \begin{bmatrix} R_1 + ree^{\mathrm{T}} - 2r_1 e^{\mathrm{T}} & R_2 + ree^{\mathrm{T}} - 2r_1 e^{\mathrm{T}} \\ R_3 + ree^{\mathrm{T}} - 2r_2 e^{\mathrm{T}} & R_4 + ree^{\mathrm{T}} - 2r_2 e^{\mathrm{T}} \end{bmatrix}.$$

Therefore, the initial minimization problem of the MSE under the constraint (6) has been reworded to the minimization of J given by (19) with respect to the $M-1$ coefficients $\tilde{a}$. The main advantage of this version of the MSE is that it incorporates the constraint (6) and does not use any heuristic technique in order to impose location-invariance. Such a technique is the normalization of the coefficients that would be derived by a direct application of the LMS algorithm to the minimization of the MSE given in (11) [23]. In (12)–(20) we have used a special treatment of the coefficient $a_{(M+1)/2}$ corresponding to the median sample of the tap-input vector for two reasons:

(a) For symmetrical noise pdfs, the adaptation of this particular coefficient does not affect the expected value of the output of the L-filter, since $\mathrm{E}[n_{((M+1)/2)}] = 0$. Indeed,

$$\mathrm{E}[y(k)] = s + a^{\mathrm{T}} \mathrm{E}[n_r(k)] = s + \sum_{i=1}^{(M-1)/2} (a_i - a_{M-i+1}) \mathrm{E}[n_{(i)}]. \tag{21}$$

(b) It is the most natural choice because of the symmetry of the problem conditions and the resulting mathematical tractability.

The filter coefficient for the median sample $x_{((M+1)/2)}$ is obtained by using

$$a_{(M+1)/2} = 1 - [e^{\mathrm{T}} \mid e^{\mathrm{T}}] \tilde{a}. \tag{22}$$

The steepest descent algorithm for the minimization of the MSE in (19) is formulated as follows:

$$\tilde{a}(k+1) = \tilde{a}(k) + \tfrac{1}{2}\mu[-\nabla J(k)], \tag{23}$$

where μ is the adaptation step and $\nabla J(k)$ is the gradient of MSE with respect to $\tilde{a}(k)$ given by

$$\nabla J(k) = -2p' + (R' + R'^{\mathrm{T}}) \tilde{a}(k) = -2p' + 2R_s' \tilde{a}(k). \tag{24}$$

R_s' denotes the symmetric part of matrix R':

$$R_s' = \begin{bmatrix} R_1 + ree^{\mathrm{T}} - (r_1 e^{\mathrm{T}} + e r_1^{\mathrm{T}}) & R_2 + ree^{\mathrm{T}} - (r_1 e^{\mathrm{T}} + e r_2^{\mathrm{T}}) \\ R_2^{\mathrm{T}} + ree^{\mathrm{T}} - (r_2 e^{\mathrm{T}} + e r_1^{\mathrm{T}}) & R_4 + ree^{\mathrm{T}} - (r_2 e^{\mathrm{T}} + e r_2^{\mathrm{T}}) \end{bmatrix}. \tag{25}$$

By substituting (24) into (23) the following location-invariant steepest descent algorithm is obtained:

$$\tilde{a}(k+1) = \tilde{a}(k) + \mu[p' - R_s' \tilde{a}(k)]. \tag{26}$$

In the following, we shall drop the primes from p' and R_s' without loss of generality. The simplest way to develop an estimate of the gradient $\nabla J(k)$ is to use instantaneous estimates for p and R_s. Let $\tilde{n}_r$ denote the following $(M-1)\times 1$ vector:

$$\tilde{\boldsymbol{n}}_r=[\boldsymbol{n}_{r1}^{\mathrm{T}}\,|\,\boldsymbol{n}_{r2}^{\mathrm{T}}]^{\mathrm{T}}. \tag{27}$$

$\boldsymbol{p}$ and $\boldsymbol{R}_{\mathrm{s}}$ can be rewritten as follows:

$$\boldsymbol{p}=\mathrm{E}[n_{((M+1)/2)}(n_{((M+1)/2)}\boldsymbol{e}_{M-1}-\tilde{\boldsymbol{n}}_r)], \tag{28}$$

$$\boldsymbol{R}_{\mathrm{s}}=\mathrm{E}[(\tilde{\boldsymbol{n}}_r-n_{((M+1)/2)}\boldsymbol{e}_{M-1})(\tilde{\boldsymbol{n}}_r-n_{((M+1)/2)}\boldsymbol{e}_{M-1})^{\mathrm{T}}]. \tag{29}$$

Thus the estimates of $\boldsymbol{p}$ and $\boldsymbol{R}_{\mathrm{s}}$ are given by

$$\hat{\boldsymbol{p}}(k)=n_{((M+1)/2)}^{k}(n_{((M+1)/2)}^{k}\boldsymbol{e}_{M-1}-\tilde{\boldsymbol{n}}_r(k)), \tag{30}$$

$$\hat{\boldsymbol{R}}_{\mathrm{s}}(k)=(\tilde{\boldsymbol{n}}(k)-n_{((M+1)/2)}^{k}\boldsymbol{e}_{M-1})(\tilde{\boldsymbol{n}}_r(k)-n_{((M+1)/2)}^{k}\boldsymbol{e}_{M-1})^{\mathrm{T}}. \tag{31}$$

The unbiased estimate for the gradient vector is the following:

$$\begin{aligned}\hat{\nabla}J(k)&=-2\hat{\boldsymbol{p}}(k)+2\hat{\boldsymbol{R}}_{\mathrm{s}}(k)\hat{\boldsymbol{a}}(k)\\&=-2((n_{((M+1)/2)}^{k}\boldsymbol{e}_{M-1}-\tilde{\boldsymbol{n}}_r(k))^{\mathrm{T}}\hat{\boldsymbol{a}}(k)-n_{((M+1)/2)}^{k})(\tilde{\boldsymbol{n}}_r(k)-n_{((M+1)/2)}^{k}\boldsymbol{e}_{M-1}),\end{aligned} \tag{32}$$

where

$$\hat{\boldsymbol{a}}(k)=[a_1(k),\ldots,a_{(M-1)/2}(k),a_{(M+3)/2}(k),\ldots,a_M(k)]^{\mathrm{T}}. \tag{33}$$

By using (30)–(33), the LMS adaptation formula is written as follows:

$$\begin{aligned}\hat{\boldsymbol{a}}(k+1)&=\hat{\boldsymbol{a}}(k)+\tfrac{1}{2}\mu[-\hat{\nabla}J(k)]\\&=\hat{\boldsymbol{a}}(k)+\mu(\tilde{\boldsymbol{n}}_r(k)-n_{((M+1)/2)}^{k}\boldsymbol{e}_{M-1})((n_{((M+1)/2)}^{k}\boldsymbol{e}_{M-1}-\tilde{\boldsymbol{n}}_r(k))^{\mathrm{T}}\hat{\boldsymbol{a}}(k)-n_{((M+1)/2)}^{k}).\end{aligned} \tag{34}$$

It can easily be recognized that the last term inside parentheses is the estimation error at time instant k. Indeed,

$$\begin{aligned}&(n_{((M+1)/2)}^{k}\boldsymbol{e}_{M-1}-\tilde{\boldsymbol{n}}_r(k))^{\mathrm{T}}\hat{\boldsymbol{a}}(k)-n_{((M+1)/2)}^{k}\\&\quad=n_{((M+1)/2)}^{k}\left\{\sum_{\substack{j=1\\j\neq(M+1)/2}}^{M}a_j(k)-1\right\}-\sum_{\substack{j=1\\j\neq(M+1)/2}}^{M}a_j(k)n_{(j)}^{k}\\&\quad=s-y(k)=\varepsilon(k).\end{aligned} \tag{35}$$

Therefore the adaptation of all coefficients except $a_{(M+1)/2}$ is described by

$$\hat{\boldsymbol{a}}(k+1)=\hat{\boldsymbol{a}}(k)+\mu\,\varepsilon(k)(\tilde{\boldsymbol{x}}_r(k)-x_{((M+1)/2)}^{k}\boldsymbol{e}_{M-1}), \tag{36}$$

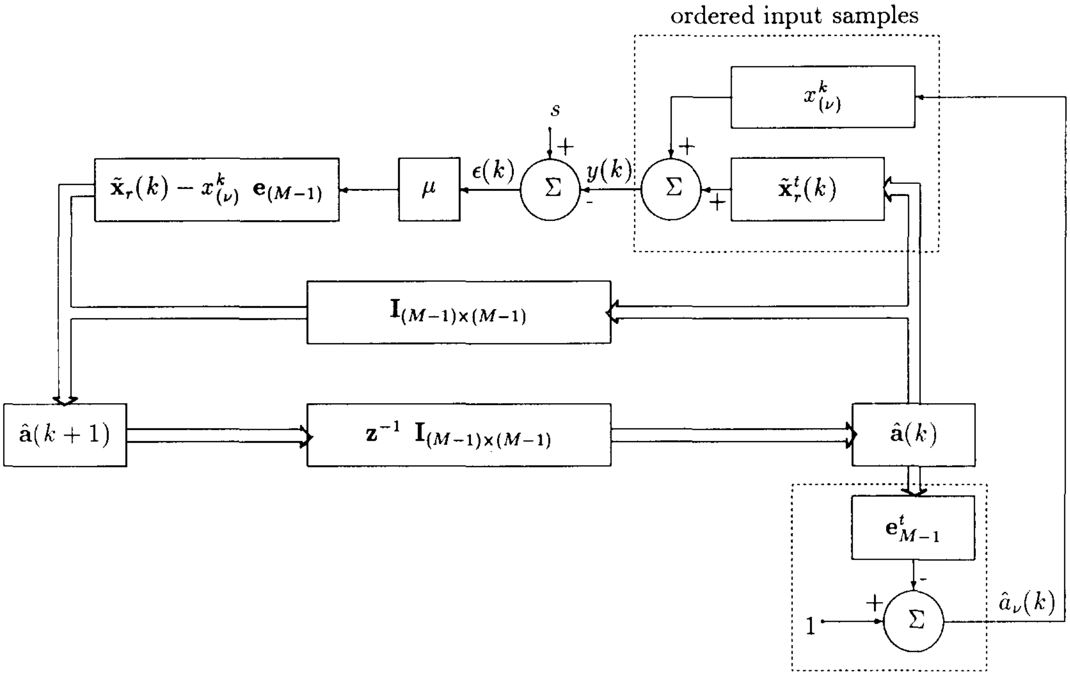


Fig. 1. Signal-flow graph representation of the location-invariant LMS L-filter algorithm. Single lines denote scalar multiplications. Double lines denote vector/matrix multiplications. $z^{-1}\boldsymbol{I}_{(M-1)\times(M-1)}$ is the transmittance matrix of a unit-delay branch. For $v=(M+1)/2$: $\hat{\boldsymbol{a}}(k)=[a_1(k),\ldots,a_{v-1}(k),\ a_{v+1}(k),\ldots,a_M(k)]^{\mathrm{T}}$, $a_v(k)$: L-filter coefficients; $\tilde{\boldsymbol{x}}_r(k)=[x^k_{(1)},\ldots,x^k_{(v-1)},\ x^k_{(v+1)},\ldots,x^k_{(M)}]^{\mathrm{T}}$, $x^k_{(v)}$: ordered input samples; $y(k)$: L-filter output; s: desired constant signal to be estimated.

where $\tilde{\boldsymbol{x}}_r(k)=[x_{(1)}(k),\ldots,x_{((M-1)/2)}(k),\ x_{((M+3)/2)}(k),\ldots,x_{(M)}(k)]^{\mathrm{T}}$ and $x_{((M+1)/2)}$ is the median input sample. The coefficient for the median sample is given by

$$\hat{a}_{(M+1)/2}(k)=1-\boldsymbol{e}^{\mathrm{T}}_{M-1}\hat{\boldsymbol{a}}(k). \tag{37}$$

The signal-flow graph representation of the location-invariant LMS adaptive L-filter algorithm is shown in Fig. 1. This graph is multidimensional in the sense that the nodes of the graph consist of vectors and that the transmittance of each branch of the graph is a scalar or a square matrix. The above-mentioned adaptation procedure uses the estimation error $\varepsilon(k)$. Therefore, the knowledge of a constant reference signal $s(k)=s$ is needed for coefficient adaptation, as it is the case for any adaptive filter. If this reference signal is not available it can be estimated from the past input samples x_i in a window $k-L\leqslant i\leqslant k$ by using a convenient estimator (e.g. the arithmetic mean). An efficient computational scheme for (35)–(37) uses $2M+2$ local variables, i.e. M variables for $\boldsymbol{x}_r(k)$, M variables for the coefficients, one variable for μ and another one for the estimation error $\varepsilon(k)$. It requires $2M$ multiplications per output sample, i.e., M multiplications for the estimated output and M multiplications for the adaptation of the coefficients except the one for the median sample. The number of the required additions per output sample is $4(M-1)$, i.e. $3(M-1)$ additions for the adaptation of the filter coefficients and $M-1$ additions for the estimated output. In addition to the total amount of multiplications and additions required the extra cost due to the sorting of M numbers has to be evaluated. However, this computational load is not large if special running ordering algorithms or structures are used. Such a simple running sorting algorithm based on the fact that at each time k the number $x_{k-1-(M-1)/2}$ is discarded from the ordered input vector $\boldsymbol{x}_r(k-1)$ and the number

$x_{k+(M-1)/2}$ is inserted in such a position in the ordered input vector, so that $\boldsymbol{x}_r(k)$ to be produced has been proposed in [20]. The number of comparisons required is $2(M-1)-\log_2 M$, i.e. a linear function of M.

3.2. Unbiased LMS adaptive L-filter

If the noise pdf is symmetrical about zero and the conditions (7) are fulfilled, the L-filter output is an unbiased estimation of the desired constant signal. Let $\boldsymbol{L}$ be the $(M-1)/2\times(M-1)/2$ exchange matrix:

$$\boldsymbol{L}=\begin{bmatrix}0 & 0 & \cdots & 0 & 1\\ 0 & 0 & \cdots & 1 & 0\\ & & \iddots & & \\ 1 & 0 & \cdots & 0 & 0\end{bmatrix}. \tag{38}$$

If the filter length M is odd, by using (7) the coefficient vector $\boldsymbol{a}$ can take the following form:

$$\boldsymbol{a}=[\boldsymbol{a}_1^{\mathrm{T}}\,|\,a_{(M+1)/2}\,|\,\boldsymbol{a}_1^{\mathrm{T}}\boldsymbol{L}]^{\mathrm{T}}. \tag{39}$$

The coefficient $a_{(M+1)/2}$ is given by

$$a_{(M+1)/2}=1-2\boldsymbol{e}^{\mathrm{T}}\boldsymbol{a}_1. \tag{40}$$

The correlation matrix is partitioned as in (16). The following form of the error function can be obtained:

$$\begin{aligned}J&=r-2[\boldsymbol{a}_1^{\mathrm{T}}\,|\,\boldsymbol{a}_1^{\mathrm{T}}]\begin{bmatrix}r\boldsymbol{e}-\boldsymbol{r}_1\\ r\boldsymbol{e}-\boldsymbol{L}\boldsymbol{r}_2\end{bmatrix}\\ &\quad+[\boldsymbol{a}_1^{\mathrm{T}}\,|\,\boldsymbol{a}_1^{\mathrm{T}}]\begin{bmatrix}\boldsymbol{R}_1+r\boldsymbol{e}\boldsymbol{e}^{\mathrm{T}}-2\boldsymbol{r}_1\boldsymbol{e}^{\mathrm{T}} & \boldsymbol{R}_2\boldsymbol{L}+r\boldsymbol{e}\boldsymbol{e}^{\mathrm{T}}-2\boldsymbol{r}_1\boldsymbol{e}^{\mathrm{T}}\\ \boldsymbol{L}\boldsymbol{R}_3+r\boldsymbol{e}\boldsymbol{e}^{\mathrm{T}}-2\boldsymbol{L}\boldsymbol{r}_2\boldsymbol{e}^{\mathrm{T}} & \boldsymbol{L}\boldsymbol{R}_4\boldsymbol{L}+r\boldsymbol{e}\boldsymbol{e}^{\mathrm{T}}-2\boldsymbol{L}\boldsymbol{r}_2\boldsymbol{e}^{\mathrm{T}}\end{bmatrix}\begin{bmatrix}\boldsymbol{a}_1\\ \boldsymbol{a}_1\end{bmatrix}.\end{aligned} \tag{41}$$

Since the noise distribution is symmetrical about zero, the correlation matrix of the ordered noise samples exhibits a double symmetry which is expressed by the following equations:

$$\boldsymbol{r}_2=\boldsymbol{L}\boldsymbol{r}_1,\quad \boldsymbol{R}_3=\boldsymbol{R}_2^{\mathrm{T}},\quad \boldsymbol{R}_1=\boldsymbol{L}\boldsymbol{R}_4\boldsymbol{L},\quad \boldsymbol{R}_2\boldsymbol{L}=\boldsymbol{L}\boldsymbol{R}_3. \tag{42}$$

By employing (42) in (41), the MSE takes the form

$$J=r-4\boldsymbol{a}_1^{\mathrm{T}}\boldsymbol{v}+2\boldsymbol{a}_1^{\mathrm{T}}\boldsymbol{R}\boldsymbol{a}_1, \tag{43}$$

where

$$\boldsymbol{v}=[r\boldsymbol{e}-\boldsymbol{r}_1]=\mathrm{E}[n_{((M+1)/2)}(n_{((M+1)/2)}\boldsymbol{e}-\boldsymbol{n}_{r1})], \tag{44}$$

$$\boldsymbol{R}=\boldsymbol{R}_1+\boldsymbol{R}_2\boldsymbol{L}+2(r\boldsymbol{e}\boldsymbol{e}^{\mathrm{T}}-2\boldsymbol{r}_1\boldsymbol{e}^{\mathrm{T}}). \tag{45}$$

Thus, the initial minimization problem of the MSE under the constraint (7) is equivalent to the minimization of (43) with respect to the $(M-1)/2$ coefficients $\boldsymbol{a}_1$. The main advantage of this version of the MSE is that it exploits the symmetry of the noise distribution and its leads to an adaptive filter whose performance characteristics (stability and convergence rate) are controlled by a matrix of dimensions $(M-1)/2\times(M-1)/2$ which has a smaller eigenvalue spread than the correlation matrix of the ordered noise samples $\boldsymbol{R}$ and the matrix $\boldsymbol{R}_s'$ of the location-invariant LMS adaptive L-filter, as it will be seen in the subsequent sections.

Again a steepest descent algorithm can be written:

$$\tilde{\boldsymbol{a}}_1(k+1) = \tilde{\boldsymbol{a}}_1(k) + \tfrac{1}{2}\mu[-\nabla J(k)]. \tag{46}$$

The gradient of the MSE with respect to the coefficients $\tilde{\boldsymbol{a}}_1(k)$is given by the following relation:

$$\nabla J(k) = -4\boldsymbol{v} + 2(R + R^{\mathrm{T}})\tilde{\boldsymbol{a}}_1(k) = -4\boldsymbol{v} + 4R_s\tilde{\boldsymbol{a}}_1(k), \tag{47}$$

where R_s is the symmetric part of matrix R given by

$$R_s = \boldsymbol{R}_1 + \boldsymbol{L}\boldsymbol{R}_2 + 2(r\boldsymbol{e}\boldsymbol{e}^{\mathrm{T}} - \boldsymbol{r}_1\boldsymbol{e}^{\mathrm{T}} - \boldsymbol{e}\boldsymbol{r}_1^{\mathrm{T}}). \tag{48}$$

Let $\boldsymbol{u}$ be the following $(M-1)/2 \times 1$ vector:

$$\boldsymbol{u} = [n_{(1)} + n_{(M)}, n_{(2)} + n_{(M-1)}, \ldots, n_{((M-1)/2)} + n_{((M+3)/2)}]^{\mathrm{T}}. \tag{49}$$

R_s can be rewritten in the form

$$R_s = \mathrm{E}[\boldsymbol{n}_{r1}\boldsymbol{u}^{\mathrm{T}} + 2n_{((M+1)/2)}(n_{((M+1)/2)}\boldsymbol{e} - \boldsymbol{n}_{r1})\boldsymbol{e}^{\mathrm{T}} - 2n_{((M+1)/2)}\boldsymbol{e}\boldsymbol{n}_{r1}^{\mathrm{T}}]. \tag{50}$$

By substituting (47) in (46) we get

$$\tilde{\boldsymbol{a}}_1(k+1) = \tilde{\boldsymbol{a}}_1(k) + 2\mu[\boldsymbol{v} - R_s\tilde{\boldsymbol{a}}_1(k)]. \tag{51}$$

The instantaneous estimates for $\boldsymbol{v}$ and R_s are derived as follows:

$$\hat{\boldsymbol{v}}(k) = n^k_{((M+1)/2)}(n^k_{((M+1)/2)}\boldsymbol{e} - \boldsymbol{n}_{r1}(k)), \tag{52}$$

$$\hat{R}_s(k) = \boldsymbol{n}_{r1}(k)\boldsymbol{u}^{\mathrm{T}}(k) + 2n^k_{((M+1)/2)}(n^k_{((M+1)/2)}\boldsymbol{e} - \boldsymbol{n}_{r1}(k))\boldsymbol{e}^{\mathrm{T}} - 2n^k_{((M+1)/2)}\boldsymbol{e}\boldsymbol{n}_{r1}^{\mathrm{T}}(k). \tag{53}$$

The unbiased estimate for the gradient vector is given by

$$\begin{aligned}\hat{\nabla} J(k) = -4\hat{\boldsymbol{v}}(k) + 4\hat{R}_s(k)\hat{\boldsymbol{a}}_1(k) = -4\{&n^k_{((M+1)/2)}(1 - 2\boldsymbol{e}^{\mathrm{T}}\hat{\boldsymbol{a}}_1(k))(n^k_{((M+1)/2)}\boldsymbol{e} - \boldsymbol{n}_{r1}(k)) \\ &+ 2n^k_{((M+1)/2)}\boldsymbol{e}\boldsymbol{n}_{r1}^{\mathrm{T}}(k)\hat{\boldsymbol{a}}_1(k) - \boldsymbol{n}_{r1}(k)\boldsymbol{u}^{\mathrm{T}}(k)\hat{\boldsymbol{a}}_1(k)\},\end{aligned} \tag{54}$$

where

$$\hat{\boldsymbol{a}}_1(k) = [a_1(k), \ldots, a_{(M-1)/2}(k)]^{\mathrm{T}}. \tag{55}$$

Let $\tilde{\boldsymbol{x}}_u(k) = 2s\boldsymbol{e} + \boldsymbol{u}(k)$. The estimation error at time instant k is given by

$$\varepsilon(k) = s - y(k) = s - \{(1 - 2\boldsymbol{e}^{\mathrm{T}}\hat{\boldsymbol{a}}_1(k))x^k_{((M+1)/2)} + \tilde{\boldsymbol{x}}_u^{\mathrm{T}}(k)\hat{\boldsymbol{a}}_1(k)\}. \tag{56}$$

Let $\boldsymbol{w}(k)$ be the following $(M-1)/2 \times 1$ vector:

$$\boldsymbol{w}(k) = [x^k_{(1)} - x^k_{(M)}, x^k_{(2)} - x^k_{(M-1)}, \ldots, x^k_{((M-1)/2)} - x^k_{((M+3)/2)}]^{\mathrm{T}} \tag{57}$$

and $v(k)$ be the following scalar:

$$v(k) = \hat{\boldsymbol{a}}_1^{\mathrm{T}}(k)\boldsymbol{w}(k). \tag{58}$$

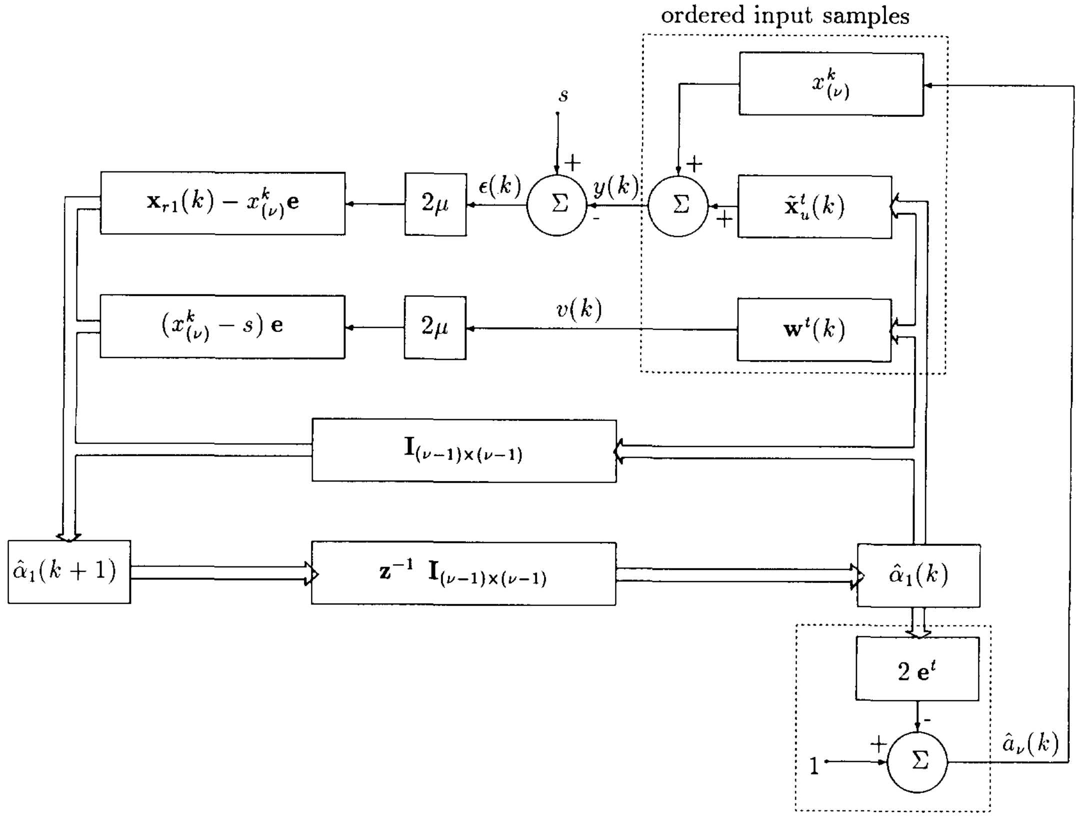


Fig. 2. Signal-flow graph representation of the unbiased LMS L-filter algorithm. Single lines denote scalar transmittances. Double lines indicate vectorial transmittances. For $v=(M+1)/2$: $\hat{\boldsymbol{a}}_1(k)=[a_1(k), \ldots, a_{(v-1)}(k)]^T$, $a_v(k)$: distinct L-filter coefficients; $\boldsymbol{x}_{r1}(k)=[x^k_{(1)}, \ldots, x^k_{(v-1)}]^T$, $x^k_{(v)}$, $\boldsymbol{x}_{r2}(k)=L[x^k_{(v+1)}, \ldots, x^k_{(M)}]^T$: ordered input samples; $\tilde{\boldsymbol{x}}_u(k)=\boldsymbol{x}_{r1}(k)+\boldsymbol{x}_{r2}(k)$, $\boldsymbol{w}(k)=\boldsymbol{x}_{r1}(k)-\boldsymbol{x}_{r2}(k)$; $y(k)$: L-filter output; s: desired constant signal to be estimated.

Then the i-th element of the gradient vector, $[\nabla J(k)]_i$ is given by

$$[\nabla J(k)]_i = -4\{\varepsilon(k)(x^k_{(i)} - x^k_{((M+1)/2)}) - v(k)(s - x^k_{((M+1)/2)})\}. \tag{59}$$

Therefore the LMS unbiased adaptation formula is described by

$$\hat{\boldsymbol{a}}_1(k+1) = \hat{\boldsymbol{a}}_1(k) + 2\mu\{\varepsilon(k)(\boldsymbol{x}_{r1}(k) - x^k_{((M+1)/2)}\boldsymbol{e}) - v(k)(s - x^k_{((M+1)/2)})\boldsymbol{e}\}, \tag{60}$$

where $\boldsymbol{x}_{r1}(k) = [x_{(1)}(k), \ldots, x_{((M-1)/2)}(k)]^T$. The signal-flow graph for the proposed unbiased LMS adaptive L-filter is shown in Fig. 2. The remark made in the case of the location-invariant LMS adaptive L-filter about the knowledge of the reference constant signal holds in this case, too. The unbiased LMS adaptive L-filter requires $(3M+7)/2$ local variables, i.e. $(M+1)/2$ variables for the filter coefficients, M variables for the ordered samples, one variable for the adaptation step μ, one for the estimation error $\varepsilon(k)$ and another one for $v(k)$. Its computational demands are $(5M-1)/2$ multiplications per output sample, i.e. $(M+1)/2$ multiplications for the estimated output, $(M-1)/2$ multiplications for the computation of $v(k)$ and $(3M-1)/2$ multiplications for the adaptation of filter coefficients and $4M-5$ additions per output sample, i.e. $(M-1)/2$ additions for the estimated output, $M-2$ additions for the computation of $v(k)$, and $2(M-1)$ additions for the adaptation of filter coefficients. The same amount of comparisons, as in the case of the location-invariant LMS L-filter, is also required.

4. Convergence properties of the proposed constrained adaptive *L*-filters

There are two objectives in the analysis of the LMS algorithm [2, 11]. First, the convergence of the mean coefficient-error vector to zero as the iteration number k tends to infinity must be proven. This type of convergence is called convergence in the mean. Second, the convergence of the ensemble average mean square error $\mathrm{E}[J(k)]$ over all samples of input data processes to a steady-state value as k approaches infinity has to be considered. This steady-state value is equal to the minimum MSE plus the excess MSE because the LMS algorithm relies on a noisy estimate for the gradient vector $\nabla J(k)$. This is the so-called convergence in the mean square.

It is known [2, 11] that the convergence in the mean is determined by the largest eigenvalue of the correlation matrix whereas the rate of convergence of the filter coefficients to the optimal ones is determined by its smallest eigenvalue when the eigenvalues of the correlation matrix are widely spread. On the contrary, when the eigenvalues of the correlation matrix are widely spread, the excess MSE produced by the LMS algorithm is primarily determined by the largest eigenvalues and the rate of convergence of $\mathrm{E}[J(k)]$ is not affected by the spread of the eigenvalues so much as the rate of convergence of the filter coefficients.

In this section, the convergence properties of the proposed constrained adaptive LMS L-filters are studied. It is shown that the extreme eigenvalues of the matrices $\boldsymbol{R}_s$ and R_s determine the convergence in the mean of the location-invarant and the unbiased LMS L-filter, respectively. The relation of the extreme eigenvalues of the matrix $\boldsymbol{R}_s$ and its trace to the extreme eigenvalues and the trace of the correlation matrix of the ordered noise samples $\boldsymbol{R}$ are also studied.

However, the proof of the above-described statements is not unconditional. In the derivation of the convergence in the mean each ordered sample noise vector $\boldsymbol{n}_r(k)$ is treated as statistically independent of all previous ordered noise sample vectors $\boldsymbol{n}_r(l)$, $l=0,\ldots,k-1$, i.e.,

$$\mathrm{E}[\boldsymbol{n}_r(k)\boldsymbol{n}_r^{\mathrm{T}}(l)]=\mathbf{0}, \quad l=0,1,\ldots,k-1. \tag{61}$$

A similar assumption called fundamental assumption plays a central role in the analysis of the adaptive linear filters [11]. The following assumption will be used in the derivation of the convergence in the mean square: the ordered noise samples are mutually Gaussian-distributed random variables for all k.

4.1. *Location-invariant adaptive LMS L-filter*

Let $\boldsymbol{c}(k)$ denote the coefficient-error vector:

$$\boldsymbol{c}(k)=\hat{\boldsymbol{a}}(k)-\tilde{\boldsymbol{a}}_{\mathrm{o}}, \tag{62}$$

where $\tilde{\boldsymbol{a}}_{\mathrm{o}}$ is the vector of the optimal coefficients except the one for the median input sample. The coefficient-error for $a_{(M+1)/2}(k)$ is given by

$$c_{(M+1)/2}(k)=-\boldsymbol{e}_{M-1}^{\mathrm{T}}\boldsymbol{c}(k). \tag{63}$$

Thus it suffices to prove the convergence in the mean for $\boldsymbol{c}(k)$, i.e. that $\mathrm{E}[\boldsymbol{c}(k)]$ tends to zero when k tends to infinity.

If the LMS algorithm is rewritten in terms of the coefficient-error vector $\boldsymbol{c}(k)$ and the independence of $\boldsymbol{c}(k)$ from $\tilde{\boldsymbol{n}}_r(k)$ is exploited, by taking the expected values of both sides and using (28)–(29) we obtain

$$\mathrm{E}[\boldsymbol{c}(k+1)]=(\boldsymbol{I}-\mu\boldsymbol{R}_s)\mathrm{E}[\boldsymbol{c}(k)]+\mu(\boldsymbol{p}-\boldsymbol{R}_s\tilde{\boldsymbol{a}}_{\mathrm{o}}), \tag{64}$$

where $\boldsymbol{I}$ is the identity matrix. It is easily recognized that the second term of the right-hand side of (64) equals zero. Consequently,

$$\mathrm{E}[\boldsymbol{c}(k+1)] = (\boldsymbol{I} - \mu \boldsymbol{R}_{\mathrm{s}})\mathrm{E}[\boldsymbol{c}(k)]. \tag{65}$$

Therefore the mean of $\boldsymbol{c}(k)$ converges to zero as k approaches infinity when the following conditions are satisfied:

$$\boldsymbol{R}_{\mathrm{s}} \text{ is positive definite,} \quad 0 < \mu < 2/\lambda_{\max}, \tag{66}$$

where $\lambda_{\max}$ is the largest eigenvalue of the matrix $\boldsymbol{R}_{\mathrm{s}}$.

Let $z(k)$ be the coefficient-error vector transformed to the principal coordinate system:

$$z(k) = \boldsymbol{Q}^{\mathrm{T}} \boldsymbol{c}(k), \tag{67}$$

where the columns of $\boldsymbol{Q}$ are the eigenvectors associated with the eigenvalues of the matrix $\boldsymbol{R}_{\mathrm{s}}$. Let also τ_i denote the time constant for the i-th element of the expected value $\mathrm{E}[z(k)]$, i.e. the time required for the amplitude of the $\mathrm{E}[z_i(k)]$ to decay to $1/e$ of its initial value $z_i(0)$. It is known [11] that the i-th time constant is given by

$$\tau_i = \frac{-1}{\ln(1 - \mu\lambda_i)} \tag{68}$$

and the transient behavior of the i-th L-filter coefficient in the mean is described by

$$\mathrm{E}[\hat{a}_i(k)] = \tilde{a}_{i,\mathrm{o}} + \sum_{j=1}^{M-1} Q_{ij} z_j(0)(1 - \mu\lambda_j)^k, \tag{69}$$

where Q_{ij} is the ij-element of the matrix $\boldsymbol{Q}$. Therefore, each coefficient except the one for the median input sample converges in the mean as a weighted sum of exponentials of the form $(1 - \mu\lambda_j)^k$. The overall time constant τ_a for the expected value of any coefficient is bounded by

$$\frac{-1}{\ln(1 - \mu\lambda_{\max})} \leqslant \tau_a \leqslant \frac{-1}{\ln(1 - \mu\lambda_{\min})}, \tag{70}$$

where $\lambda_{\min}$ is the smallest eigenvalue of the matrix $\boldsymbol{R}_{\mathrm{s}}$.

The proof of the convergence in the mean square and the derivation of bounds for the adaptation step μ is based on the existence of a recursive equation for the coefficient-error correlation matrix $\boldsymbol{\phi}(k) = \mathrm{E}[\boldsymbol{c}(k)\boldsymbol{c}^{\mathrm{T}}(k)]$. In the derivation of this equation fourth-order moments of the order statistics must be evaluated. Such a computation is possible only numerically. Accordingly, the decomposition of the fourth-order moments of the order statistics in terms of smaller-order moments and especially in second-order moments has to be pursued. This is only possible for mutually Gaussian random variables. Therefore we shall invoke the Gaussian assumption which has been used extensively in all attempts to analyze the convergence of the ensemble-average MSE [7, 8, 11]. It is possible to use analysis similar to that of [11] to prove that a necessary condition for $\mathrm{E}[J(k)]$ to be convergent is

$$0 < \mu < 2/\mathrm{tr}[\boldsymbol{R}_{\mathrm{s}}], \tag{71}$$

where tr[ ] stands for the trace of the bracketed matrix. Use of (71) guarantees also the convergence in the mean. A tighter bound, three times smaller than that of (71), is proposed in [7].

It is worthwhile to compare the performance of the location-invariant adaptive LMS L-filter with the one of the adaptive LMS L-filters proposed in [18, 21, 23]. To do so, the extreme eigenvalues of $\boldsymbol{R}_s$ have to be related to the extreme eigenvalues of $\boldsymbol{R}$. Let us suppose that the eigenvalues of an $M \times M$ matrix $\boldsymbol{A}$ have been arranged according to their magnitude. We use the convention

$$\lambda_{(1)}(\boldsymbol{A}) \leqslant \lambda_{(2)}(\boldsymbol{A}) \leqslant \cdots \leqslant \lambda_{(M)}(\boldsymbol{A}), \tag{72}$$

i.e. $\lambda_{(1)}(\boldsymbol{A})$ is the smallest eigenvalue of $\boldsymbol{A}$, $\lambda_{(M)}(\boldsymbol{A})$ is the largest eigenvalue of $\boldsymbol{A}$. We obtain the following proposition.

Proposition. (*a*) *The smallest eigenvalue of* $\boldsymbol{R}_s$, $\lambda_{(1)}(\boldsymbol{R}_s)$, *is related to the second in magnitude eigenvalue of* $\boldsymbol{R}$, $\lambda_{(2)}(\boldsymbol{R})$, *by the inequality*

$$\lambda_{(1)}(\boldsymbol{R}_s) \leqslant \lambda_{(2)}(\boldsymbol{R}) + 3\mathrm{E}[n^2], \tag{73}$$

where $\mathrm{E}[n^2]$ *is the mean square value of the noise.*
(*b*) *The largest eigenvalue of* $\boldsymbol{R}_s$, $\lambda_{(M-1)}(\boldsymbol{R}_s)$, *is related to the largest eigenvalue of* $\boldsymbol{R}$, $\lambda_{(M)}(\boldsymbol{R})$ *as follows*:

$$\lambda_{(M-1)}(\boldsymbol{R}_s) \leqslant \lambda_{(M)}(\boldsymbol{R}) + 3\mathrm{E}[n^2]. \tag{74}$$

The proof of the above proposition is given in Appendix A.

By using (73) the following lower bound on the slowest rate of convergence of the location-invariant adaptive LMS L-filter in terms of the second in magnitude eigenvalue of $\boldsymbol{R}$ can be derived:

$$\tau_{\max} = \frac{-1}{\ln(1-\mu\lambda_{(1)}(\boldsymbol{R}_s))} \geqslant \frac{-1}{\ln(1-\mu(\lambda_{(2)}(\boldsymbol{R}) + 3\mathrm{E}[n^2]))}. \tag{75}$$

The inequality (74) implies that the location-invariant LMS L-filter is convergent in the mean if the adaptation step is bounded by

$$0 < \mu < \frac{2}{\lambda_{(M)}(\boldsymbol{R}) + 3\mathrm{E}[n^2]}. \tag{76}$$

Since $\lambda_{(M)}(\boldsymbol{R}) < \mathrm{tr}[\boldsymbol{R}] = M\mathrm{E}[n^2]$ a more conservative bound would be

$$0 < \mu < \frac{2}{(M+3)\mathrm{E}[n^2]}. \tag{77}$$

It can easily be shown that the bound on the adaptation step (77) guarantees also convergence in the mean square. Indeed,

$$\mathrm{tr}[\boldsymbol{R}_s] = \mathrm{tr}[\boldsymbol{R}] + (M-2)\mathrm{E}[n^2_{((M+1)/2)}] - 2\sum_{i=1}^{M} \mathrm{E}[n_{(i)}n_{((M+1)/2)}]. \tag{78}$$

If the last term (i.e. the sum term) is non-negative, then

$$\mathrm{tr}[\boldsymbol{R}_s] \leqslant \mathrm{tr}[\boldsymbol{R}] + (M-2)\mathrm{E}[n^2_{((M+1)/2)}] \leqslant \mathrm{tr}[\boldsymbol{R}] + 3\mathrm{E}[n^2] = (M+3)\mathrm{E}[n^2]. \tag{79}$$

Therefore a smaller adaptation step than the one that would be used for an adaptive LMS L-filter based on the correlation matrix of the ordered noise samples $\boldsymbol{R}$ is appropriate.

4.2. Unbiased adaptive LMS L-filter

Using the same reasoning it can be proven that the necessary and sufficient conditions for the mean coefficient vector $E[\hat{a}_1(k)]$ to be convergent are

$$\text{the matrix } R_s \text{ to be positive definite,} \quad 0<\mu<1/\lambda_{\max}, \tag{80}$$

where $\lambda_{\max}$ is the largest eigenvalue of the matrix R_s. A necessary condition for the ensemble average $E[J(k)]$ to be convergent is

$$0<\mu<1/\mathrm{tr}[R_s]. \tag{81}$$

5. Numerical study and simulation examples

The MSE is expressed as a quadratic form of the correlation matrix of ordered noise samples $\boldsymbol{R}$ in (11). Furthermore, an effort has been made to relate the extreme eigenvalues of the matrix which controls the adaptation procedure in the case of the location-invariant adaptive LMS L-filter to the eigenvalues of $\boldsymbol{R}$. Thus it is helpful to compute the eigenvalue spread of the matrices $\boldsymbol{R}$ for several L-filter lengths M and for the uniform, Gaussian and Laplacian noise distributions. To do so, the Householder reduction to a symmetric tridiagonal matrix and the QR algorithm with implicit shifts have been applied to $\boldsymbol{R}$ [3, 28]. Several indicative values of the eigenvalue spread are given in Table 1. It is seen that the eigenvalue spread is increased with the increase of the L-filter length. For the same L-filter length the eigenvalue spread is increased as the noise distribution becomes more long-tailed.

The validity of the assumptions (A.14) and (A.22) can be deduced by inspection of Table 2 for the cases of the uniform and Gaussian noise when $M=5$ and for the Laplacian noise when $M=9$.

The smallest eigenvalue of the matrices that have been involved in the derivation of the proposed adaptive filter structures has been evaluated for several M in the case of Gaussian noise. The results are given in

Table 1

Eigenvalue spread ($\lambda_{\max}/\lambda_{\min}$) of the correlation matrix of the ordered noise samples

M	Uniform noise $E[n^2]=1$	Gaussian noise $E[n^2]=1$	Laplacian noise $E[n^2]=1$
3	10.242639	10.560249	11.214899
5	47.036057	57.845813	74.734245
7	127.001450	172.546034	254.631378
9	266.162070	384.774761	973.757474

Table 2

Second in order and largest eigenvalue of $\boldsymbol{R}$ and $\boldsymbol{P}_{s,M}$

Type of noise	$\lambda_{(2)}(\boldsymbol{R})$	$\lambda_{(2)}(\boldsymbol{P}_{s,M})$	$\lambda_{(M)}(\boldsymbol{R})$	$\lambda_{(M)}(\boldsymbol{P}_{s,M})$
Uniform ($M=5$)	0.113349	0.086904	3.600937	3.522601
Gaussian ($M=5$)	0.108597	0.108597	3.621358	3.621358
Laplacian ($M=9$)	0.022675	0.018969	7.140832	7.140832

Table 3
Smallest eigenvalue of $\boldsymbol{R}$, $\boldsymbol{R}_s$ and R_s for the Gaussian noise distribution

M	Correlation matrix		Location-invariant	Unbiased
	$\lambda_{(1)}(\boldsymbol{R})$	$\lambda_{(2)}(\boldsymbol{R})$	$\boldsymbol{R}_s$	R_s
5	0.062604	0.108597	0.108597	0.146244
7	0.031849	0.046911	0.046911	0.057162
9	0.019249	0.025896	0.025896	0.029656
11	0.012885	0.016381	0.016381	0.018092

Table 3. The second in magnitude eigenvalue of the correlation matrix of the ordered noise samples $\boldsymbol{R}$ has also been included in the same table for comparison purposes. The smallest eigenvalues of the matrices that control the adaptation procedure both in the location-invariant L-filter and the unbiased L-filter are larger than that of the matrix $\boldsymbol{R}$. Furthermore, it is seen that the smallest eigenvalue of $\boldsymbol{R}_s$ equals $\lambda_{(2)}(\boldsymbol{R})$ for all L-filter lengths. It is obvious that inequality (73) is satisfied.

A similar attempt has been made for the largest eigenvalue of these matrices. From the inspection of Table 4 it is observed that the largest eigenvalue of the matrix which affects the stability of the location-invariant adaptive LMS L-filter is the same with that of $\boldsymbol{R}$, whereas the maximal eigenvalue of the matrix which affects the stability of the unbiased adaptive LMS L-filter is much smaller than that of the correlation matrix of ordered noise samples. Therefore, if the choice of the adaptation step is made by using (77), the stability of the proposed location-invariant adaptive LMS L-filter is guaranteed. It is readily verified that the inequality (74) is satisfied for all L-filter lengths.

The time constants (68) for the elements of the mean transformed coefficient-error vector $\mathrm{E}[z(k)]$ are given in Table 5 when an L-filter of length 5 is used for the estimation of a constant signal corrupted by Gaussian additive white noise both for standard LMS and for location-invariant LMS. The adaptation step used is $\mu = 0.001$. The smallest eigenvalue of $\boldsymbol{R}_s$ is larger than the smallest eigenvalue of $\boldsymbol{R}$. Therefore, the upper bound for the rate of convergence is reduced. We can deduce that the overall time constant for

Table 4
Largest eigenvalue of $\boldsymbol{R}$, $\boldsymbol{R}_s$ and R_s for the Gaussian noise distribution

M	Correlation matrix $\boldsymbol{R}$	Location-invariant $\boldsymbol{R}_s$	Unbiased R_s
5	3.621358	3.621358	0.697468
7	5.495397	5.495397	0.732133
9	7.406547	7.406547	0.752009
11	9.338802	9.338802	0.764901

Table 5
Time constants corresponding to the eigenvalues of $\boldsymbol{R}$ and $\boldsymbol{R}_s$

$\lambda_i(\boldsymbol{R})$	τ_i	$\lambda_i(\boldsymbol{R}_s)$	τ_i
0.062604	15973	0.108597	9208
0.108597	9208	0.109112	9165
0.207441	4820	0.595101	1680
1.0	1000	3.621358	276
3.621358	276	—	—

the location-invariant adaptive L-filter coefficients is smaller than the corresponding overall time constant of the standard LMS adaptive L-filter coefficients.

The proposed constrained adaptive LMS L-filters have been implemented by using C language and have been tested for one-dimensional signals for uniform, Gaussian and Laplacian distributions. It is known [22] that the optimal L-filter for the uniform noise is the midpoint:

$$a_1 = a_M = 0.5, \quad a_2 = \cdots = a_{M-1} = 0, \tag{82}$$

whereas for the Gaussian noise it is the arithmetic mean:

$$a_i = \frac{1}{M}, \quad i = 1, \ldots, M. \tag{83}$$

In the experiments with uniform and Gaussian noise we shall use an L-filter with 5 coefficients (i.e., $M = 5$). The optimal L-filter for the Laplacian distribution is given by (9). We shall consider the optimal L-filter coefficients reported in [5] for $M = 9$:

$$a_1 = a_9 = -0.01899, \quad a_2 = a_8 = 0.02904, \quad a_3 = a_7 = 0.06965, \quad a_4 = a_6 = 0.23795, \quad a_5 = 0.3646. \tag{84}$$

As measures for convergence, we shall consider either the coefficient estimation error $\Lambda(\boldsymbol{a}, k)$ evaluated over a single realization of an experiment, which is defined as follows:

$$\Lambda(\boldsymbol{a}, k) = \frac{1}{M} \sum_{j=1}^{M} (a_j(k) - a_{j,o})^2, \tag{85}$$

where $a_{j,o}$, $j = 1, \ldots, M$ are the optimal coefficients or the ensemble average of the squared coefficient-error $E[(a_j(k) - a_{j,o})^2]$ over 200 independent trials of an experiment for all j. A more detailed explanation on the need of ensemble averaging can be found in [11].

First of all, the noise reduction capability (i.e., the ratio of output noise power to input noise power) of the proposed constrained adaptive LMS L-filters is demonstrated in Table 6 for zero mean uniform, Gaussian and Laplacian noise distributions. The initial L-filter and the choice for the adaptation step in each case are described below. It is seen that the proposed adaptive nonlinear filters behave better for Laplacian additive white noise.

In the following set of experiments the performance of the location-invariant adaptive LMS L-filter has been measured in a variety of input noise distributions.

Table 6

Noise reduction capability of the proposed constrained adaptive LMS L-filters

L-filter type	Noise distribution	Variance	M	Noise reduction (dB)
Location invariant	uniform	0.083	5	−8.489
	Gaussian	1.0	5	−6.961
	Laplacian	2.0	9	−11.573
Unbiased	Laplacian	2.0	9	−10.971

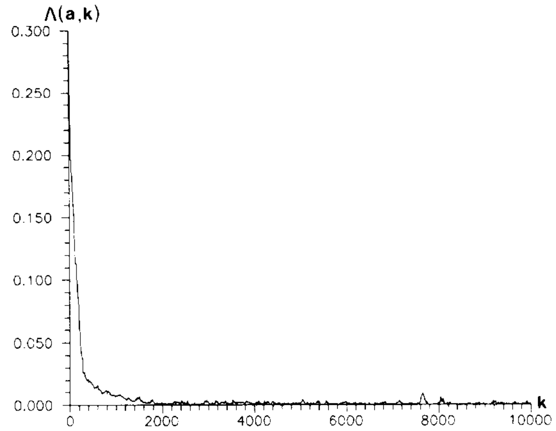


Fig. 3a. Coefficient estimation error of the location-invariant adaptive LMS L-filter for uniform noise, when the initial filter is median ($M=5$).

Next, the case of uniform white noise having zero mean and variance 0.083 is described. The initial L-filter is the median filter, i.e., the worst initial choice with respect to the optimal L-filter (i.e., midpoint) in order to test the convergence behavior of the proposed adaptive L-filter. The adaptation step is $\mu=0.1$. The matrix $\boldsymbol{Q}$ which diagonalizes $\boldsymbol{R}_s$ is found by numerical methods. By replacing the eigenvalues λ_i and the matrix $\boldsymbol{Q}$ it is obtained that the overall time constant for both $E[a_1(k)]$ and $E[a_5(k)]$ is 23 iterations. The coefficient estimation error $\Lambda(\boldsymbol{a}, k)$ is shown in Fig. 3a and the plot of ensemble average of the squared coefficient-error for a_1 is given in Fig. 3b. The plot of the filter coefficients is shown in Figs. 4a and 4b. The convergence of the coefficients to the optimal ones is obvious. Although we have not imposed initially any symmetry on coefficients, we observe that the coefficients exhibit a symmetry about the median.

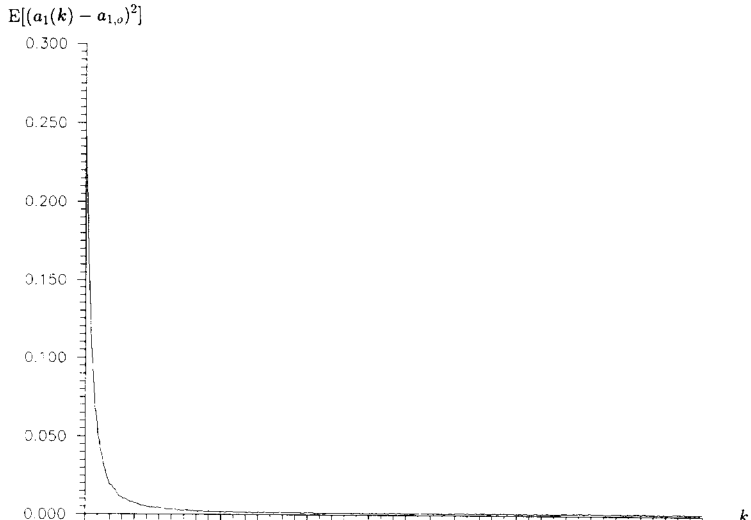
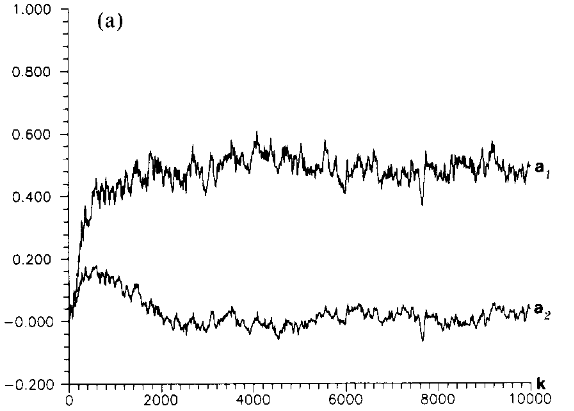


Fig. 3b. Ensemble average of the squared coefficient-error for a_1, $E[\{a_1(k)-0.5\}^2]$, vs. the iteration index k of the location-invariant adaptive LMS L-filter for uniform noise, when the initial filter is median ($M=5$).



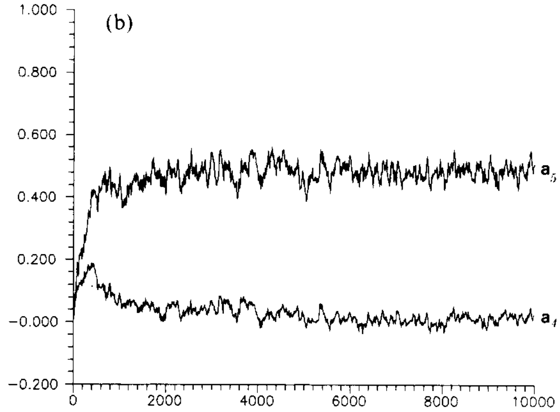


Fig. 4. Coefficient convergence of the location-invariant adaptive LMS L-filter for uniform noise, when the initial filter is median ($M=5$). (a) Coefficients a_1, a_2. (b) Coefficients a_4, a_5.

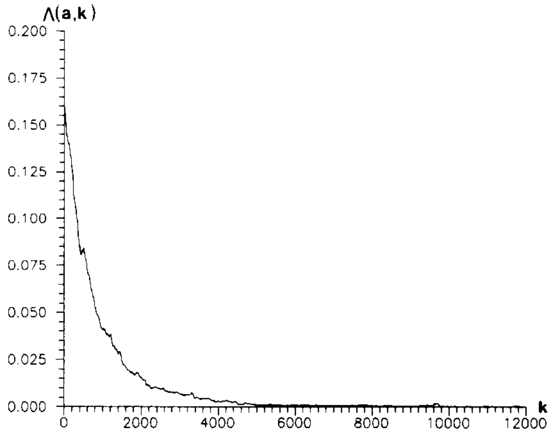


Fig. 5a. Coefficient estimation error of the location-invariant adaptive LMS L-filter for Gaussian noise, when the initial filter is median ($M=5$).

When the input constant signal is corrupted by Gaussian white noise having zero mean and unit variance the initial filter choice is again the median filter. The adaptation step is chosen as $\mu=0.001$. From (69), the overall time constant for both $\mathrm{E}[a_1(k)]$ and $\mathrm{E}[a_5(k)]$ is 1337 iterations and the overall time constant for both $\mathrm{E}[a_2(k)]$ and $\mathrm{E}[a_4(k)]$ is 3369, respectively. It is verified by a single realization of the experiment that the overall time constant for a_1 and a_5 is 1500 and for a_2 and a_4 is 2200 iterations, respectively, in this experiment. The overall time constant for a_3 is found to be 1500 iterations. Although the independence assumption used in the theoretical analysis is rather strong, it can be seen that it provides reasonable bounds on the overall time constants. The coefficient estimation error $\Lambda(\boldsymbol{a}, k)$ is plotted in Fig. 5a and the ensemble average of the squared coefficient-error for a_3 corresponding to median is shown in Fig. 5b.

When the input constant signal is corrupted by Laplacian white noise having zero mean and variance 2.0, the initial L-filter is chosen to be midpoint (82). The choice for the adaptation step is $\mu=0.003$. The ensemble average of the $\varepsilon^2(k)$ over 200 independent trials of the experiment is shown in Fig. 6. This curve is an approximation to the ensemble averaged learning curve of the location-invariant adaptive LMS L-filter for the initial choices made for the L-filter coefficients and the adaptation step. From (69), the overall

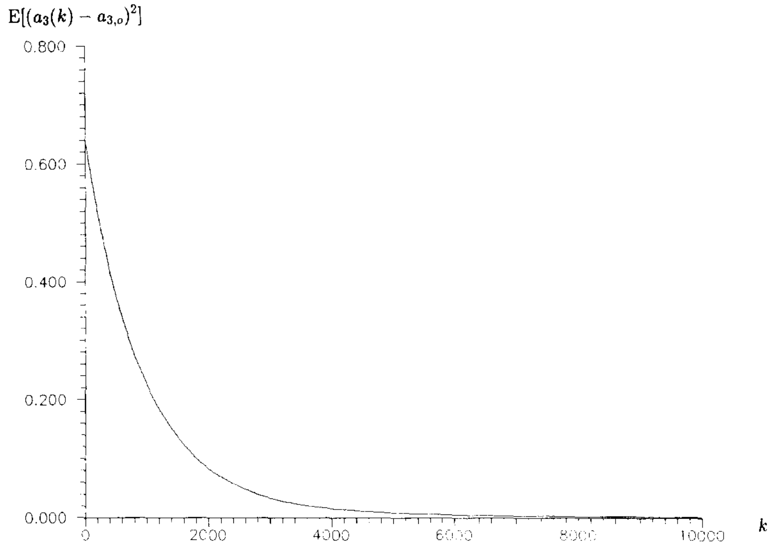


Fig. 5b. Ensemble average of the squared coefficient-error for a_3, $E[\{a_3(k)-0.2\}^2]$, versus the iteration index k of the location-invariant adaptive LMS L-filter for Gaussian noise, when the initial filter is median ($M=5$).

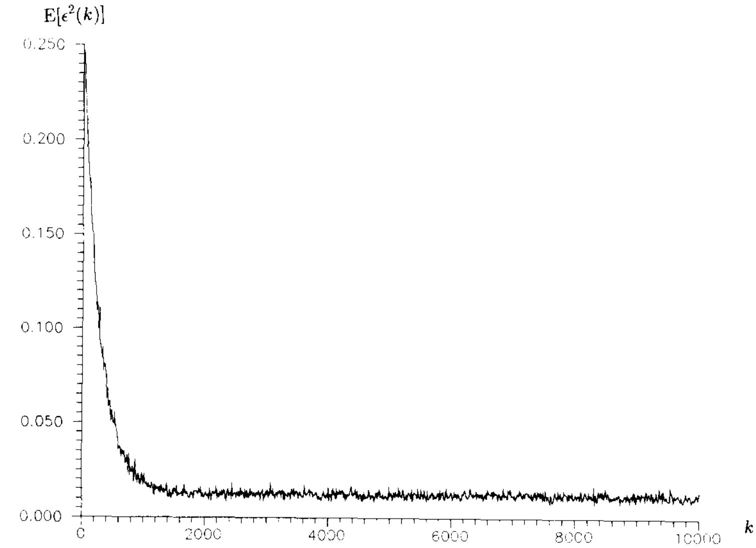


Fig. 6. Ensemble average of the squared estimation error $\varepsilon^2(k)$ versus the iteration index k of the location-invariant LMS L-filter for Laplacian noise, when the initial filter is midpoint ($M=9$).

time constant for both $E[a_1(k)]$ and $E[a_9(k)]$ is 276 iterations. It is found that the overall time constant for a_1 and a_9 is 100 iterations in this simulation. The coefficient estimation error $\Lambda(\boldsymbol{a}, k)$ is plotted in Fig. 7. It is seen that the coefficient estimation error converges to a small constant value close to zero after about 6000 iterations. The slow convergence rate is explained by the fact that the eigenvalue spread for the correlation matrix of the ordered noise samples is increased as the noise distribution becomes more long-tailed and as the length of the L-filter increases.

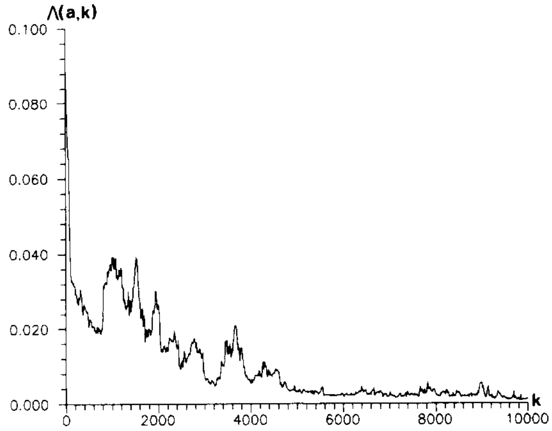


Fig. 7. Coefficient estimation error of the location-invariant adaptive LMS L-filter for Laplacian noise, when the initial filter is midpoint ($M=9$).

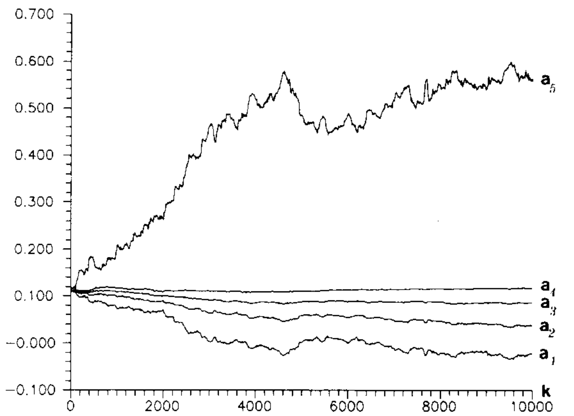


Fig. 8. Coefficient convergence of the unbiased adaptive LMS L-filter for Laplacian noise, when the initial filter is arithmetic mean ($M=9$).

The performance of the unbiased LMS L-filter has been measured in the case of white Laplacian noise. The initial L-filter is chosen to be arithmetic mean and $\mu=0.0001$. The plot of coefficients is shown in Fig. 8. Although the adaptation step is 30 times smaller than the one used in the location-invariant adaptive LMS L-filter, a faster rate of convergence is observed. In general it has been observed that the unbiased adaptive LMS L-filter performs better when the noise distribution is more long-tailed.

6. Conclusions

Two novel nonlinear adaptive LMS L-filters are presented which have the ability to incorporate the constraints underlying the location-invariant and the unbiased estimation. The algorithms to update the filter coefficients have been derived. The convergence in the mean and in the mean square of the proposed adaptive LMS L-filters has been discussed under the independence assumption. Comparison of the rate of convergence between the proposed location-invariant adaptive LMS L-filter and the adaptive LMS L-filters which are based on the correlation matrix of the ordered noise samples has been made. It has been verified that the time constant corresponding to the smallest eigenvalue of the matrix which controls the performance of the location-invariant adaptive LMS L-filter is smaller than the time constant corresponding to the smallest eigenvalue of the correlation matrix of the ordered noise samples which controls the performance of the standard adaptive LMS L-filter. The proposed filters can easily adapt to the noise probability distribution. They can be used for filtering of both long-tailed and short-tailed distributions. It has been found that the independence theory can provide useful bounds on the rate of convergence of the L-filter coefficients. However, the role of the independence theory is an open question yet even in the case of linear adaptive filtering. An analysis similar to that presented in [12, 15] is subject of ongoing research.

Appendix A

We shall use similarity transformations to prove inequalities (73), (74). It is well known that the eigenvalues remain unchanged under similarity transformations. A similarity transformation $\boldsymbol{K}^{-1}\boldsymbol{R}\boldsymbol{K}$ of $\boldsymbol{R}$ is obtained by subtracting multiples of the i-th column of $\boldsymbol{R}$ from each of the other columns and then adding the same multiples of all other rows to the i-th row [28]. This similarity transformation is expressed mathematically by means of an $M \times M$ matrix $\boldsymbol{K}$ of the form

$$\boldsymbol{K}=\begin{bmatrix} 1 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 \\ 0 & 1 & & 0 & 0 & 0 & & 0 \\ \vdots & & & & & & & \\ 0 & 0 & & 1 & 0 & 0 & & 0 \\ -k_{i1} & -k_{i2} & & -k_{i\,i-1} & 1 & -k_{i\,i+1} & & k_{iM} \\ 0 & 0 & & 0 & 0 & 1 & & 0 \\ \vdots & & & & & & & \vdots \\ 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 1 \end{bmatrix}. \tag{A.1}$$

The inverse of matrix $\boldsymbol{K}$ is obtained if the sign of the elements of the i-th row, $k_{ij}, j \neq (M+1)/2$ is reversed. We are interested in the case

$$k_{(M+1)/2\,j}=2, \quad j \neq (M+1)/2. \tag{A.2}$$

Let $\boldsymbol{P}_1$ denote the result of the similarity transformation $\boldsymbol{K}^{-1}\boldsymbol{R}\boldsymbol{K}$:

$$\boldsymbol{P}_1=\boldsymbol{R}+\begin{bmatrix} -2r_{1v} & \cdots & -2r_{1v} & 0 & -2r_{1v} & \cdots & -2r_{1v} \\ -2r_{2v} & \cdots & -2r_{2v} & 0 & -2r_{2v} & \cdots & -2r_{2v} \\ \vdots & & & & & & \vdots \\ -2r_{v-1\,v} & \cdots & -2r_{v-1\,v} & 0 & -2r_{v-1\,v} & \cdots & -2r_{v-1\,v} \\ * & & * & * & * & & * \\ -2r_{v+1\,v} & & -2r_{v+1\,v} & 0 & -2r_{v+1\,v} & \cdots & -2r_{v+1\,v} \\ \vdots & & & & & & \vdots \\ -2r_{Mv} & \cdots & -2r_{Mv} & 0 & -2r_{Mv} & \cdots & -2r_{Mv} \end{bmatrix}, \tag{A.3}$$

where v denotes the index $(M+1)/2$ and $*$ denotes 'don't care' elements. It is known [28] that the interchange of rows i and j and the corresponding columns i and j of matrix $\boldsymbol{P}_1$ is a similarity transformation. The matrix $\boldsymbol{K}_{ij}$ of this transformation is equal to the identity matrix except in rows i and j, which are of the form

$$\begin{array}{ccc} & \text{col. } i & \text{col. } j \\ \text{row } i & \multicolumn{2}{c}{\multirow{2}{*}{$\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$}} \\ \text{row } j & & \end{array}. \tag{A.4}$$

For the matrices $\boldsymbol{K}_{ij}$ of the above form (A.4) we have $\boldsymbol{K}_{ij}^{-1}=\boldsymbol{K}_{ij}$. Let us denote by $\boldsymbol{A} \propto \boldsymbol{B}$ the relation of $\boldsymbol{A}$ and $\boldsymbol{B}$ by a similarity transformation. Then

$$\begin{aligned} \boldsymbol{P}_1 &\propto \boldsymbol{K}_{(M+1)/2,(M+3)/2}\boldsymbol{P}_1\boldsymbol{K}_{(M+1)/2,(M+3)/2} \\ &\propto \boldsymbol{K}_{(M+3)/2,(M+5)/2}\boldsymbol{K}_{(M+1)/2,(M+3)/2}\boldsymbol{P}_1\boldsymbol{K}_{(M+1)/2,(M+3)/2}\boldsymbol{K}_{(M+3)/2,(M+5)/2} \\ &\propto \cdots \propto \boldsymbol{K}_{M-1,M}\cdots\boldsymbol{K}_{(M+1)/2,(M+3)/2}\boldsymbol{P}_1\boldsymbol{K}_{(M+1)/2,(M+3)/2}\cdots\boldsymbol{K}_{M-1,M}=\boldsymbol{P}_2. \end{aligned} \tag{A.5}$$

The resulting matrix $\boldsymbol{P}_2$ is given by

$$\boldsymbol{P}_2=\begin{bmatrix} r_{11}-2r_{1v} & \cdots & r_{1\,v-1}-2r_{1v} & r_{1\,v+1}-2r_{1v} & \cdots & r_{1M}-2r_{1v} & r_{1v} \\ r_{21}-2r_{2v} & \cdots & r_{2\,v-1}-2r_{2v} & r_{2\,v+1}-2r_{2v} & \cdots & r_{2M}-2r_{2v} & r_{2v} \\ \vdots & & & & & & \vdots \\ r_{M1}-2r_{Mv} & \cdots & r_{M\,v-1}-2r_{Mv} & r_{M\,v+1}-2r_{Mv} & \cdots & r_{MM}-2r_{Mv} & r_{Mv} \\ * & & * & * & & * & * \end{bmatrix}. \tag{A.6}$$

Since matrices $\boldsymbol{P}_2$ and $\boldsymbol{R}$ are related by a similarity transformation, the following equation holds:

$$\lambda_{(i)}(\boldsymbol{P}_2)=\lambda_{(i)}(\boldsymbol{R}). \tag{A.7}$$

Let $\boldsymbol{P}_\mathrm{s}$ denote the symmetric part of $\boldsymbol{P}_2$. It is given by

$$\boldsymbol{P}_\mathrm{s}=\begin{bmatrix} r_{11}-2r_{1v} & r_{12}-r_{1v}-r_{2v} & \cdots & r_{1M}-r_{1v}-r_{Mv} & * \\ r_{12}-r_{1v}-r_{2v} & r_{22}-2r_{2v} & \cdots & r_{2M}-r_{2v}-r_{Mv} & * \\ \vdots & & & \vdots & \\ r_{1M}-r_{1v}-r_{Mv} & r_{2M}-r_{2v}-r_{Mv} & \cdots & r_{MM}-2r_{Mv} & * \\ * & * & \cdots & * & * \end{bmatrix}. \tag{A.8}$$

Let us denote by $\boldsymbol{P}_{\mathrm{s},M-1}$ the following matrix:

$$\boldsymbol{P}_{\mathrm{s},M-1}=\{p_{s\,ij}\}, \quad i,j=1,\ldots,M-1. \tag{A.9}$$

Then the following inequalities hold [3]:

$$\lambda_{(1)}(\boldsymbol{P}_\mathrm{s})\leqslant\lambda_{(1)}(\boldsymbol{P}_{\mathrm{s},M-1})\leqslant\lambda_{(2)}(\boldsymbol{P}_\mathrm{s}), \qquad \lambda_{(M-1)}(\boldsymbol{P}_\mathrm{s})\leqslant\lambda_{(M-1)}(\boldsymbol{P}_{\mathrm{s},M-1})\leqslant\lambda_{(M)}(\boldsymbol{P}_\mathrm{s}). \tag{A.10}$$

It is easily recognized that the matrix $\boldsymbol{R}_\mathrm{s}$ is related to $\boldsymbol{P}_{\mathrm{s},M-1}$ as follows:

$$\boldsymbol{R}_\mathrm{s}=\boldsymbol{P}_{\mathrm{s},M-1}+\mathrm{E}[n^2_{((M+1)/2)}]\boldsymbol{e}_{M-1}\boldsymbol{e}^\mathrm{T}_{M-1}. \tag{A.11}$$

The second matrix of the right-hand side of (A.11) is semidefinite with $\lambda_{(1)}=\cdots=\lambda_{(M-2)}=0$ and $\lambda_{(M-1)}=(M-1)\mathrm{E}[n^2_{((M+1)/2)}]$. It is obvious that

$$\lambda_{(1)}(\boldsymbol{R}_\mathrm{s})\leqslant\lambda_{(1)}(\boldsymbol{P}_{\mathrm{s},M-1})+(M-1)\mathrm{E}[n^2_{((M+1)/2)}]. \tag{A.12}$$

By using (A.10) the following is obtained:

$$\lambda_{(1)}(\boldsymbol{R}_\mathrm{s})\leqslant\lambda_{(2)}(\boldsymbol{P}_{\mathrm{s},M})+(M-1)\mathrm{E}[n^2_{((M+1)/2)}]. \tag{A.13}$$

If the following inequality holds:

$$\lambda_{(2)}(\boldsymbol{P}_{\mathrm{s},M})\leqslant\lambda_{(2)}(\boldsymbol{R}), \tag{A.14}$$

then

$$\lambda_{(1)}(\boldsymbol{R}_\mathrm{s})\leqslant\lambda_{(2)}(\boldsymbol{R})+(M-1)\mathrm{E}[n^2_{((M+1)/2)}]. \tag{A.15}$$

The assumption (A.14) is valid in all simulations performed for any type of noise and for different filter lengths as can be seen in Section 5. The mean square value of the median for the uniform distribution is given by [6]

$$\mathrm{E}[n^2_{((M+1)/2)}] = \frac{3}{M+2}\,\mathrm{E}[n^2]. \tag{A.16}$$

For the Gaussian distribution an approximate formula is found in [13]

$$\mathrm{E}[n^2_{((M+1)/2)}] \simeq \frac{\pi/2}{M-1+\pi/2}\,\mathrm{E}[n^2], \tag{A.17}$$

whereas for Laplacian distribution the following approximate formula is proposed in [14]:

$$\mathrm{E}[n^2_{((M+1)/2)}] \simeq \frac{2/3}{M-5/4}\,\mathrm{E}[n^2]. \tag{A.18}$$

Since $3(M-5/4) > M+2 > M-1+\pi/2$ we obtain

$$\lambda_{(1)}(\boldsymbol{R}_s) \leqslant \lambda_{(2)}(\boldsymbol{R}) + \frac{M-1}{M-1+\pi/2}\,3\mathrm{E}[n^2] \leqslant \lambda_{(2)}(\boldsymbol{R}) + 3\mathrm{E}[n^2], \tag{A.19}$$

which is inequality (73).

For the largest eigenvalue the following inequality holds:

$$\lambda_{(M-1)}(\boldsymbol{R}_s) \leqslant \lambda_{(M-1)}(\boldsymbol{P}_{s,M-1}) + (M-1)\mathrm{E}[n^2_{((M+1)/2)}]. \tag{A.20}$$

By using (A.10) the following is obtained:

$$\lambda_{(M-1)}(\boldsymbol{R}_s) \leqslant \lambda_{(M)}(\boldsymbol{P}_{s,M}) + (M-1)\mathrm{E}[n^2_{((M+1)/2)}]. \tag{A.21}$$

If the following inequality holds:

$$\lambda_{(M)}(\boldsymbol{P}_{s,M}) \leqslant \lambda_{(M)}(\boldsymbol{R}), \tag{A.22}$$

then

$$\begin{aligned}\lambda_{(M-1)}(\boldsymbol{R}_s) \leqslant \lambda_{(M)}(\boldsymbol{R}) + (M-1)\mathrm{E}[n^2_{((M+1)/2)}] &\leqslant \lambda_{(M)}(\boldsymbol{R}) + \frac{M-1}{M-1+\pi/2}\,3\mathrm{E}[n^2] \\ &\leqslant \lambda_{(M)}(\boldsymbol{R}) + 3\mathrm{E}[n^2].\end{aligned} \tag{A.23}$$

The assumption (A.22) was valid in all simulations performed, as can be seen in Section 5.

References

[1] O. Aggazi, D.G. Messerschmitt and D.A. Hodges, "Nonlinear echo cancellation", *IEEE Trans. Comm.*, Vol. COM-30, No. 11, November 1982, pp. 2421–2433.

[2] M. Bellanger, *Adaptive Digital Filters and Signal Analysis*, Marcel Dekker, New York, 1987.

[3] R. Bellman, *Introduction to Matrix Analysis*, McGraw-Hill, New York, 1960.

[4] C.G. Boncelet Jr., "Algorithms to compute order statistic distributions", *SIAM J. Sci. Statist. Comput.*, Vol. 8, No. 5, September 1987.

[5] A. Bovik, T.S. Huang and D.C. Munson, "A generalization of median filtering using linear combinations of order statistics", *IEEE Trans. Acoust. Speech Signal Process.*, Vol. ASSP-31, No. 6, December 1983, pp. 1342–1349.

[6] H.A. David, *Order Statistics*, Wiley, New York, 1981.
[7] A. Feuer and E. Weinstein, "Convergence analysis of LMS filters with uncorrelated Gaussian data", *IEEE Trans. Acoust. Speech Signal Process.*, Vol. ASSP-33, No. 1, February 1985, pp. 222–230.
[8] W.A. Gardner, "Learning characteristics of stochastic-gradient-descent algorithms: A general study, analysis, and critique", *Signal Processing*, Vol. 6, No. 2, April 1984, pp. 113–133.
[9] Z. Govindarajulu, "On moments of order statistics and quasi-ranges from normal populations", *Ann. Math. Statist.*, Vol. 34, 1962, pp. 633–651.
[10] T.I. Haweel and P.M. Clarkson, "Analysis and generalization of a median adaptive filter", *Proc. Internat. Conf. Acoust. Speech Signal Process.*, Albuquerque, 1990, pp. 1269–1272.
[11] S. Haykin, *Adaptive Filter Theory*, Prentice Hall, Englewood Cliffs, NJ, 1986.
[12] S. Jones, R. Cavin III and W. Reed, "Analysis of error-gradient adaptive linear estimators for a class of stationary dependent processes", *IEEE Trans. Inform. Theory*, Vol. IT-28, No. 2, March 1982, pp. 318–329.
[13] B.I. Justusson, "Median filtering: Statistical properties", in: T.S. Huang, ed., *Two-Dimensional Digital Signal Processing II, Transforms and Median Filters*, Springer, New York, 1981.
[14] Y.-H. Lee and S. Tantaratana, "A simple formula for the variance of the median filter output with Laplacian input", *IEEE Trans. Signal Process.*, Vol. 39, No. 4, April 1991, pp. 984–986.
[15] O. Macchi and E. Eweda, "Second-order convergence analysis of stochastic adaptive linear filtering", *IEEE Trans. Automat. Control*, Vol. AC-28, No. 1, January 1983, pp. 76–85.
[16] F. Palmieri, "A backpropagation algorithm for multilayer hybrid order statistic filters", *Proc. Internat. Conf. Acoust. Speech Signal Process.*, Glasgow, England, 1989, pp. 1179–1182.
[17] F. Palmieri, "Adaptive recursive order statistic filters", *Proc. Internat. Conf. Acoust. Speech Signal Process.*, Albuquerque, 1990, pp. 1229–1232.
[18] F. Palmieri and C.G. Boncelet Jr., "A class of nonlinear adaptive filters", *Proc. Internat. Conf. Acoust. Speech Signal Process.*, New York, 1988, pp. 1483–1486.
[19] F. Palmieri and C.G. Boncelet Jr., "A new class of order statistic filters", *IEEE Trans. Acoust. Speech Signal Process.*, Vol. ASSP-37, No. 5, May 1989, pp. 691–701.
[20] I. Pitas, "Fast algorithms for running ordering and max/min selection algorithms", *IEEE Trans. Circuits and Systems*, Vol. CAS-36, No. 6, June 1989, pp. 795–804.
[21] I. Pitas and A.N. Venetsanopoulos, "Adaptive *L*-filters", *Proc. European Conf. Circuit Theory and Design*, Brighton, England, 1989.
[22] I. Pitas and A.N. Venetsanopoulos, *Nonlinear Digital Filters: Principles and Applications*, Kluwer Academic Publishers, Dordrecht, 1990.
[23] I. Pitas and A.N. Venetsanopoulos, "Adaptive filters based on order statistics", *IEEE Trans. Signal Process.*, Vol. ASSP-39, No. 2, February 1991, pp. 518–522.
[24] K. Saarineen and Y. Neuvo, "Adaptation of weighted median filters", *Proc. V European Signal Processing Conference*, Barcelona, Spain, 1990, pp. 269–272.
[25] K.S. Shrikantan, "Recurrence relations between the pdf's of order statistic and some applications", *Ann. Math. Statist.*, Vol. 27, 1962, pp. 169–177.
[26] G.L. Sicuranza and G. Ramboni, "Adaptive nonlinear digital filters using distributed arithmetic", *IEEE Trans. Acoust. Speech Signal Process.*, Vol. ASSP-34, No. 3, June 1986, pp. 518–526.
[27] D. Teichroew, "Tables of expected values of order statistics and products of order statistics for samples of size twenty and less from the normal distribution", *Ann. Math. Statist.*, Vol. 27, 1956, pp. 364–387.
[28] J.H. Wilkinson, *The Algebraic Eigenvalue Problem*, Oxford Univ. Press, London, 1965.