CONSTRAINED ADAPTIVE LMS L-FILTERS

C. Kotropoulos, I. Pitas

Department of Electrical Engineering
University of Thessaloniki
Thessaloniki 540 06, GREECE

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 properties of the proposed filters are considered. Both of them can adapt well to a variety of noise probability distributions ranging from the short-tailed ones to long-tailed ones.

1. Introduction

Adaptive filters constitute an important part of statistical signal processing. They have been applied in a wide variety of problems [1]. An effort has been attempted to combine the adaptive filtering and the nonlinear filtering [2]. Extentions of the LMS and RLS algorithms have been proposed in [3]. Adaptive hybrid filter structures have been proposed in [4].

The main purpose of this paper is to extend the standard LMS algorithm by applying it to the adaptation of the coefficients of the L-filters in order to incorporate constraints imposed on the coefficients. L-filters are defined as linear combinations of the ordered data in the filter window, i.e. the output of the L-filter at time instant k is given by:

$$y(k) = \sum_{j=1}^{M} a_j x_{(j)}^{k} \quad (1)$$

where $x_{(j)}^{k}$ is the i-th largest observed data and M is assumed to be odd. It has been proven [5,6] that the optimal L-filter for estimating a constant signal in the presence of additive white noise should be either location-invariant or unbiased. Let $\mathbf{e}_M$ denotes the M×1 unitary vector, i.e., $\mathbf{e}_M = [1,\dots,1]^T$ and $\mathbf{a} = [a_1,\dots,a_M]^T$ denotes the vector of L-filter coefficients. The necessary and sufficient condition for a location-invariant L-filter is:

$$\mathbf{e}_M^T \mathbf{a} = 1 \quad (2)$$

The sufficient conditions for an unbiased L-filter are:

$$\mathbf{e}_M^T \mathbf{a} = 1$$
$$a_j = a_{M-j+1} \quad j=1,\dots,(M-1)/2 \quad (3)$$

and the noise distribution should be symmetric about zero.

Two novel schemes are derived by rewriting the normal equations in a form that takes into account the constraints and by using instantaneous values for the correlations of the ordered noise samples in order to derive an estimate for the gradient vector.

2. Constrained LMS adaptive L-filters

A constant signal s corrupted by zero-mean additive white noise is considered. Thus, the input samples have the form $x_i = s + n_i$, where n_i are i.i.d. random variables having zero mean. The noise distribution is assumed symmetric about zero.

First, the adaptation formula for the location- invariant adaptive LMS L-filter is derived. Let $\mathbf{n}_r$ denote the vector of the ordered noise samples, i.e.:

$$\mathbf{n}_r = [n_{(1)}, n_{(2)}, \dots, n_{(M)}]^T \quad (4)$$

The mean squared error J is given by:

$$J = E[\,(y(k)-s)^2\,] = \mathbf{a}^T \mathbf{R}\, \mathbf{a} \quad (5)$$

where $\mathbf{R}$ is the correlation matrix of the ordered noise samples. Its (i,j) element is given by $r_{ij} = E[n_{(i)}n_{(j)}]$, i,j =1,...,M. Let $\mathbf{e}$ denotes the (M-1)/2×1 unitary vector. The coefficient vector $\mathbf{a}$ is rewritten as:

$$\mathbf{a} = [\boldsymbol{\alpha}_1^T \mid 1-\mathbf{e}^T\boldsymbol{\alpha}_1-\mathbf{e}^T\boldsymbol{\alpha}_2 \mid \boldsymbol{\alpha}_2^T]^T \qquad (6)$$

where $\boldsymbol{\alpha}_1=[a_1,\ldots,a_{(M-1)/2}]^T$ and $\boldsymbol{\alpha}_2=[a_{(M+3)/2},\ldots,a_M]^T$. Similarly, $\mathbf{n}_r$ can be partitioned in the form (6), i.e.:

$$\mathbf{n}_r = [\mathbf{n}_{r1}^T \mid n_{((M+1)/2)} \mid \mathbf{n}_{r2}^T]^T \qquad (7)$$

where $n_{((M+1)/2)}$ is the median noise sample, $\mathbf{n}_{r1}=[n_{(1)},\ldots,n_{((M-1)/2)}]^T$ and $\mathbf{n}_{r2}=[n_{((M+3)/2)},\ldots,n_{(M)}]^T$. Then, the correlation matrix $\mathbf{R}$ is partitioned as follows:

$$\mathbf{R}= \begin{bmatrix} \mathbf{R}_1 & \mathbf{r}_1 & \mathbf{R}_2 \\ \mathbf{r}_1^T & r & \mathbf{r}_2^T \\ \mathbf{R}_3 & \mathbf{r}_2 & \mathbf{R}_4 \end{bmatrix} \qquad (8)$$

where $\mathbf{R}_1= E[\mathbf{n}_{r1}\mathbf{n}_{r1}^T]$, $\mathbf{R}_2= E[\mathbf{n}_{r1}\mathbf{n}_{r2}^T]$, $\mathbf{R}_3= \mathbf{R}_2^T$, $\mathbf{R}_4= E[\mathbf{n}_{r2}\mathbf{n}_{r2}^T]$, $r= E[n^2_{((M+1)/2)}]$, $\mathbf{r}_1= E[n_{((M+1)/2)}\mathbf{n}_{r1}]$ and $\mathbf{r}_2= E[n_{((M+1)/2)}\mathbf{n}_{r2}]$. The MSE (5) is rewritten as:

$$J = r - 2\,\tilde{\mathbf{a}}^T\,\mathbf{p}' + \tilde{\mathbf{a}}^T\,\mathbf{R}'\,\tilde{\mathbf{a}} \qquad (9)$$

where

$$\tilde{\mathbf{a}} = [\boldsymbol{\alpha}_1^T \mid \boldsymbol{\alpha}_2^T]^T$$

$$\mathbf{p}' = [r\mathbf{e}^T-\mathbf{r}_1^T \mid r\mathbf{e}^T-\mathbf{r}_2^T]^T \qquad (10)$$

$$\mathbf{R}'= \begin{bmatrix} \mathbf{R}_1 + r\mathbf{e}\mathbf{e}^T - 2\mathbf{r}_1\mathbf{e}^T & \mathbf{R}_2 + r\mathbf{e}\mathbf{e}^T - 2\mathbf{r}_1\mathbf{e}^T \\ \mathbf{R}_3 + r\mathbf{e}\mathbf{e}^T - 2\mathbf{r}_2\mathbf{e}^T & \mathbf{R}_4 + r\mathbf{e}\mathbf{e}^T - 2\mathbf{r}_2\mathbf{e}^T \end{bmatrix}$$

The main advantage of this version of the MSE is that it leads to an adaptation scheme which does not use any heuristic technique to impose location invariance e.g. 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 (5). The steepest descent algorithm for the minimization of J in (9) is given by:

$$\tilde{\mathbf{a}}(k+1) = \tilde{\mathbf{a}}(k) + \mu\,[\mathbf{p}' - \mathbf{R}_s'\,\tilde{\mathbf{a}}(k)] \qquad (11)$$

where μ is the adaptation step and $\mathbf{R}_s'$ is the symmetric part of the matrix $\mathbf{R}'$. The bracketed term is the gradient $\nabla J(k)$ of MSE with respect to $\tilde{\mathbf{a}}(k)$. In the following, we shall drop the primes from $\mathbf{p}'$ and $\mathbf{R}_s'$. The simplest way to develop an estimate of the gradient $\nabla J(k)$ is to use instantaneous estimates for $\mathbf{p}$ and $\mathbf{R}_s$:

$$\hat{\mathbf{p}}(k) = n^k_{((M+1)/2)}\{\, n^k_{((M+1)/2)}\mathbf{e}_{M-1} - \tilde{\mathbf{n}}_r(k)\,\} \qquad (12)$$

$$\hat{\mathbf{R}}_s(k) = \tilde{\mathbf{n}}_r(k)\,\tilde{\mathbf{n}}_r^T(k) - n^k_{((M+1)/2)}\mathbf{e}_{M-1}\tilde{\mathbf{n}}_r^T(k) + $$
$$+ n^k_{((M+1)/2)}\{\, n^k_{((M+1)/2)}\mathbf{e}_{M-1} - \tilde{\mathbf{n}}_r(k)\,\}\,\mathbf{e}_{M-1}^T \qquad (13)$$

The LMS adaptation formula is written as follows:

$$\hat{\mathbf{a}}(k+1) = \hat{\mathbf{a}}(k) + \mu\,[\hat{\mathbf{p}}(k) - \hat{\mathbf{R}}_s(k)\,\hat{\mathbf{a}}(k)] =$$
$$= \hat{\mathbf{a}}(k) + \mu\,\varepsilon(k)\,(\,\tilde{\mathbf{x}}_r(k) - x^k_{((M+1)/2)}\mathbf{e}_{M-1}) \qquad (14)$$

where $\varepsilon(k) = s - y(k)$, $\tilde{\mathbf{x}}_r(k) = s\,\mathbf{e}_{M-1} + \tilde{\mathbf{n}}_r(k)$ and $x_{((M+1)/2)} = s + n_{((M+1)/2)}$. The coefficient for the median sample is given by:

$$\hat{a}_{(M+1)/2}(k) = 1 - \mathbf{e}_{M-1}^T\,\hat{\mathbf{a}}(k) \qquad (15)$$

We proceed next to the derivation of the unbiased LMS adaptive L-filter. Let $\mathbf{L}$ be the following $[(M-1)/2 \times (M-1)/2]$ matrix:

$$\mathbf{L} = \begin{bmatrix} 0 & \ldots & 0 & 1 \\ 0 & \ldots & 1 & 0 \\ \vdots & & & \vdots \\ 1 & \ldots & 0 & 0 \end{bmatrix} \qquad (16)$$

By using (3), the coefficient vector takes the following form:

$$\mathbf{a} = [\boldsymbol{\alpha}_1^T \mid a_{(M+1)/2} \mid \boldsymbol{\alpha}_1^T\,\mathbf{L}]^T \qquad (17)$$

where $a_{(M+1)/2}= 1 - 2\,\mathbf{e}^T\boldsymbol{\alpha}_1$. The correlation matrix of the ordered noise samples exhibits a double symmetry which is expressed by the equations:

$$\mathbf{r}_2 = \mathbf{L}\,\mathbf{r}_1,\ \mathbf{R}_3 = \mathbf{R}_2^T,\ \mathbf{R}_1 = \mathbf{L}\,\mathbf{R}_4\,\mathbf{L},\ \mathbf{R}_2\mathbf{L} = \mathbf{L}\,\mathbf{R}_3 \qquad (18)$$

By employing (18), the MSE is given by the following expression:

$$J = r - 4\,\boldsymbol{\alpha}_1^T\,\mathcal{D} + 2\,\boldsymbol{\alpha}_1^T\,\mathcal{R}\,\boldsymbol{\alpha}_1 \qquad (19)$$

where

$$\mathcal{D} = r\mathbf{e} - \mathbf{r}_1 \qquad (20)$$

$$\mathcal{R} = \mathbf{R}_1 + \mathbf{R}_2\mathbf{L} + 2\,\{\, r\,\mathbf{e}\mathbf{e}^T - 2\,\mathbf{r}_1\mathbf{e}^T\,\} \qquad (21)$$

Again, a steepest descent algorithm can be written:

$$\tilde{\mathbf{a}}_1(k+1) = \tilde{\mathbf{a}}_1(k) + 2\mu \{ \mathcal{D} - \mathcal{R}_s \tilde{\mathbf{a}}_1(k) \} \qquad (22)$$

where $\mathcal{R}_s$ is the symmetric part of matrix $\mathcal{R}$. The instantaneous estimates for $\mathcal{D}$ and $\mathcal{R}_s$ can be derived as previously. Let $\mathbf{w}(k)=[n^k_{(1)}-n^k_{(M)},\ldots, n^k_{((M-1)/2)}-n^k_{((M+3)/2)}]^T$ and $v(k) = \hat{\mathbf{a}}^T_1(k)\ \mathbf{w}(k)$. After some algebraic manipulation the following unbiased LMS adaptation formula is obtained:

$$\hat{\mathbf{a}}_1(k+1) = \hat{\mathbf{a}}_1(k) + 2\mu\{\varepsilon(k)\ [\hat{\mathbf{n}}_{r1}(k)-n^k_{((M+1)/2)}\mathbf{e}] +$$
$$+ v(k)\ n^k_{((M+1)/2)}\ \mathbf{e}\ \} \qquad (23)$$

3. Convergence properties of the proposed adaptive L-filters

In this section , the convergence in the mean of the location-invariant adaptive LMS L-filter by using the fundamental assumption [1] is proven. The proof of the convergence for the unbiased adaptive LMS L-filter is similar.

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

$$\mathbf{c}(k) = \hat{\mathbf{a}}(k) - \tilde{\mathbf{a}}_o \qquad (24)$$

The coefficient error for $a_{(M+1)/2}$ is $c_{(M+1)/2} = -\mathbf{e}^T_{M-1}\ \mathbf{c}\ (k)$. If the LMS algorithm is rewritten in terms of $\mathbf{c}(k)$ and the expected values are taken then by using the independence of the coefficient vector from the previous ordered sample vectors the following equation is obtained:

$$E[\mathbf{c}(k+1)] = (\ \mathbf{I}\ -\mu\ \mathbf{R}'_s\ )\ E[\mathbf{c}(k)] \qquad (25)$$

Therefore the mean of $\mathbf{c}(k)$ converges to zero as k tends to ∞, when $\mathbf{R}'_s$ is positive definite and the the following inequality is satisfied:

$$0 < \mu < \frac{2}{\lambda_{max}} \qquad (26)$$

where λ_{max} is the largest eigenvalue of the matrix $\mathbf{R}'_s$. Since $\mathbf{R}'_s$ is characterized by a large eigenvalue spread, the rate of convergence of the location-invariant LMS adaptive L-filter determined by its smallest eigenvalue.

The relation of the extreme eigenvalues of the matrices $\mathbf{R}_s$ and $\mathcal{R}_s$ with the eigenvalues of the correlation matrix of the ordered noise samples has been studied. Several indicative values of the eigenvalue spread of the matrix $\mathbf{R}$ found by numerical methods for various L-filter lengths and for the uniform, Gaussian and Laplacian noise distributions are given in TABLE 1. It can be 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 becomes more long-tailed. It has been observed that the smallest eigenvalue of both $\mathbf{R}_s$ and $\mathcal{R}_s$ is always larger than that of the matrix $\mathbf{R}$. This implies that the proposed LMS adaptive L-filters exhibit a faster rate of convergence than an LMS adaptive L-filter which is controlled by the matrix $\mathbf{R}$. It has also been observed that the largest eigenvalue of $\mathbf{R}_s$ is the same with that of $\mathbf{R}$ whereas the largest eigenvalue of $\mathcal{R}_s$ is much smaller than that of the correlation matrix of the ordered noise samples.

Although the independence assumption used above is rather strong, it provides reasonable bounds on the overall time constant τ_a for the coefficients of the L-filter [1]:

$$\frac{-1}{\ln(1-\mu\lambda_{max})} \le \tau_a \le \frac{-1}{\ln(1-\mu\lambda_{min})} \qquad (27)$$

where λ_{min} is the minimal eigenvalue of the matrix which controls the adaptation procedure. In the case of the location-invariant L-filter of length M=5 and for Gaussian noise, substituting λ_{min}=0.108597 and λ_{max}=3.621358 into (27) it is obtained:

$$276 \le \tau_a \le 9208 \quad \text{iterations} \qquad (28)$$

It has been verified by simulations that the overall time constant for the coefficients a_1, a_3 and a_5 is about 1000 iterations and for a_2 and a_4 2000 iterations respectively.

4. Simulation Examples

The proposed LMS constrained adaptive L-filters have been implemented using C language and have been tested for one dimensional signals for uniform, Gaussian and Laplacian distributions. As a measure for convergence, we shall consider the coefficient estimation error $\Lambda(\mathbf{a},k)$, which is defined by:

$$\Lambda(\mathbf{a},k) = \frac{1}{M} \sum_{j=1}^{M} (\ a_j(k) - a_{j,o}\ )^2 \qquad (29)$$

where $a_{j,o}$, j=1,...,M are the optimal coefficients for uniform, Gaussian and Laplacian distribution. It has been shown that the proposed nonlinear adaptive filters can adapt well to a variety of noise probability distributions ranging from the short-tailed ones (e.g. uniform) to the long-tailed ones (e.g. Laplacian).

The performance of the location-invariant LMS L-filter is shown in Figure 1. The convergence of the L-filter coefficients to the midpoint ($a_1=a_5=0.5$, $a_2=\ldots=a_4=0$) in the case of uniform white noise having zero mean and variance 0.083 is shown in Figure 1a. The initial filter is chosen to be median. The adaptation step is $\mu=0.1$. The convergence of the L-filter coefficients to the arithmetic mean ($a_1=\ldots=a_5=0.2$) in the case of Gaussian white noise having zero mean and unity variance is shown in Figure 1b. The initial filter is again median. The adaptation step is chosen as $\mu=0.001$. The convergence of the L-filter coefficients to the optimal filter reported in [5] in the case of Laplacian white noise having zero mean and variance 2.0 is shown in Figure 1c. The initial filter is midpoint and the adaptation step is $\mu=0.003$.

Acknowledgements

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

References

[1] S. Haykin, Adaptive filter theory, Prentice-Hall, Englewood Cliffs, 1986.

[2] I. Pitas, A.N. Venetsanopoulos, Nonlinear digital filters:Principles and applications, Kluwer Academic, 1990.

[3] I. Pitas, A.N. Venetsanopoulos, "Adaptive filters based on order statistics", IEEE Transactions on ASSP, in press.

[4] F. Palmieri, "A backpropagation algorithm for multilayer hybrid order statistic filters", Proc. of the Int. Conf. on ASSP, Glasgow, England, May 1989.

[5] A. Bovik, T.S. Huang, D.C. Munson, "A generalization of median filtering using linear combinations of order statistics", IEEE Transactions on ASSP, Vol. ASSP-31, No. 6, December 1983.

[6] F.Palmieri, C.G. Boncelet Jr.,"A new class of order statistic filters", Transactions on ASSP, vol. ASSP-37, No. 5, pp. 691-701, May 1989.

TABLE 1
Eigenvalue spread of the correlation matrix 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	10.560259	11.214899
5	47.036057	57.845813	74.734245
7	127.001450	172.546034	254.631378
9	266.162070	384.774761	973.757474

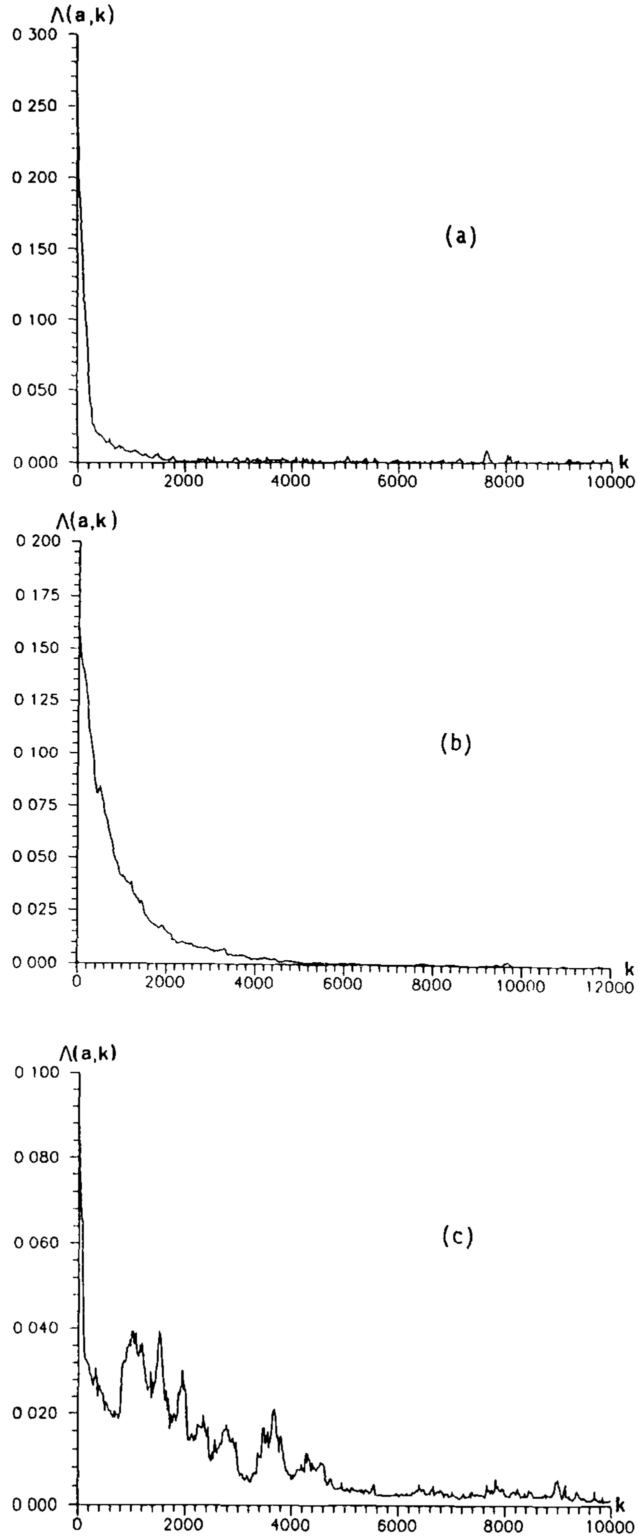

Figure 1:(a) Coefficient estimation error of the location-invariant LMS L-filter for uniform noise, when the initial filter is median (M=5).
(b) Coefficient estimation error of the location-invariant LMS L-filter for Gaussian noise, when the initial filter is median (M=5).
(c) Coefficient estimation error of the location-invariant LMS L-filter for Laplacian noise, when the initial filter is midpoint (M=9).