ADAPTIVE MULTICHANNEL L FILTERS BASED ON REDUCED ORDERING

N. Nikolaidis† I. Pitas‡

†Dept. of ECE, University of Thessaloniki, Thessaloniki, GREECE, nikolaid@zeus.csd.auth.gr
‡Dept. of Informatics, University of Thessaloniki, Thessaloniki, GREECE, pitas@zeus.csd.auth.gr

ABSTRACT

Multichannel L filters that are based on the reduced ordering principle have been proposed lately as an effective nonlinear filtering structure for multivariate data. The evaluation of the optimal coefficients for these filters requires a priori information on the signal statistics which might not be always available. To overcome this, we propose adaptive multichannel L filters that are based on the LMS algorithm. Convergence issues for the new adaptive filter structures are studied. Experiments involving color images prove the superior performance of the proposed filters in noise removal.

1 INTRODUCTION

Single channel nonlinear filtering techniques based on data ordering have been used with significant success in numerous applications, especially in the area of digital image processing. This success motivated many researchers to investigate possible extensions of order statistics filters for handling multichannel signals. The main difficulty in introducing such extensions is that unlike scalar ordering, ordering of vector-valued observations is not uniquely defined. A number of multivariate ordering techniques (marginal, partial, reduced, conditional ordering) have been proposed in the statistical literature. Among them, reduced ordering (R-ordering) seems to be the one that is the most "natural" for vector-valued observations. Furthermore, it is easy to implement and it can provide clues about outlying observations. According to R-ordering, p-channel observations $\mathbf{x}_1, \ldots, \mathbf{x}_N$, $\mathbf{x}_i = [x_{i1}, \ldots, x_{ip}]^T$ are ordered with respect to their distance d_i from some reference vector $\overline{\mathbf{a}}$. The i-th order statistic $\mathbf{x}_{[i]}$ according to R-ordering is the input vector whose distance from the reference vector is ranked i-th among distances $d_1, \ldots, d_N$. Any multivariate estimator of location (e.g. the arithmetic mean or the vector median) evaluated over a subset of the input data can be used as reference vector. However, the choice of an appropriate reference vector is crucial for the reduced ordering scheme.

A new class of optimal multichannel L filters that are based on R-ordering has been recently introduced [1],[2]. Simulations have proven that these filters perform better than other multichannel filters. The evaluation of the optimal filter coefficients requires a priori knowledge of the correlation matrices and the mean vectors for all ordered input vectors. It is therefore obvious that these filters cannot be used in situations where such information is either not available or requires extensive numerical calculations. Adaptive algorithms have been used successfully to cope with similar problems in many scientific disciplines. Among the various adaptation schemes that have been proposed, the Least Mean Squares (LMS) algorithm [3] has gained noticeable popularity mainly due to its implementation simplicity. In this paper we use a modification of this algorithm to derive equations for the adaptation of the coefficients of multichannel L filters that are based on reduced ordering.

The paper is organized as follows. Section 2 gives an overview of the optimal multichannel L filters that are based on R-ordering. The derivation of the adaptation formulas for their LMS adaptive counterparts is presented in Section 3. Convergence issues are studied in Section 4. Section 5 presents simulations involving filtering of noisy color images.

2 OPTIMAL MULTICHANNEL L FILTERS

A p-channel L filter of length N that is based on R-ordering is defined by the following equation :

$$\mathbf{y}(k) = \tilde{\mathbf{a}}^T(k)\tilde{\mathbf{x}}^T(k) \tag{1}$$

where $\tilde{\mathbf{x}}(k)$ is the $(1 \times Np)$-dimensional vector of the ordered input vectors that lie within the filter window :

$$\tilde{\mathbf{x}}(k) = \left[\begin{array}{ccc} \mathbf{x}_{[1]}^T(k), & \ldots & , \mathbf{x}_{[N]}^T(k) \end{array} \right] \tag{2}$$

and $\tilde{\mathbf{a}}(k)$ is a $(Np \times p)$ coefficient matrix. For the derivation of the optimal coefficients the input signal is assumed to be a constant vector-valued signal $\mathbf{s} = [s_1, \ldots, s_p]^T$ corrupted by additive zero-mean p-channel noise $\mathbf{n}(k) = [n_1(k), \ldots, n_p(k)]^T$.

If no constraint is imposed on the filter coefficients the filter is called an *unconstrained* multichannel L filter. It has been proven [1] that the calculation of the optimal (with respect to the output MSE) unconstrained filter coefficients requires a priori knowledge on the noise-free signal $\mathbf{s}$ as well as on the correlation matrix and the mean vector of $\tilde{\mathbf{x}}(k)$.

A multichannel L filter is said to be *location invariant* if an input of the form $\mathbf{x}'(k) = \mathbf{x}(k)+\mathbf{c}$ produces an output $\mathbf{y}'(k) = G[\mathbf{x}'(k)] = \mathbf{y}(k) + \mathbf{c}$ where $\mathbf{y}(k) = G[\mathbf{x}(k)]$. This property implies that the following constraints hold for the filter coefficients :

$$\sum_{l=0}^{N-1} [\tilde{a}(k)]_{p\cdot l+i,i} = 1, \quad i = 1,\dots,p$$
$$\sum_{l=0}^{N-1} [\tilde{a}(k)]_{p\cdot l+j,i} = 0, \quad i,j = 1,\dots,p,\ i \neq j \qquad (3)$$

where $[\tilde{a}(k)]_{m,n}$ denotes the m,n element of $\tilde{\mathbf{a}}(k)$. The above constraints can be alternatively expressed in matrix notation :

$$\tilde{\mathbf{b}}(k)\mathbf{T}^T = \mathbf{I}_p \qquad (4)$$

where $\mathbf{I}_p$ is the $(p \times p)$-dimensional identity matrix, $\mathbf{T}$ is a $(p \times Np)$-dimensional matrix given by :

$$\mathbf{T} = \begin{bmatrix} \mathbf{e}_1 & \mathbf{e}_0 & \dots & \mathbf{e}_0 \\ \ddots & \ddots & & \ddots \\ \mathbf{e}_0 & \mathbf{e}_0 & \dots & \mathbf{e}_1 \end{bmatrix} \qquad (5)$$

and the $(1 \times N)$-dimensional vectors $\mathbf{e}_1$, $\mathbf{e}_0$ are given by $\mathbf{e}_1 = [\ 1 \ \dots \ 1\ ]$, $\mathbf{e}_0 = [\ 0 \ \dots \ 0\ ]$. Finally, $\tilde{\mathbf{b}}(k)$ is a rearranged form of $\tilde{\mathbf{a}}(k)$ given by :

$$\tilde{\mathbf{b}}(k) = \tilde{\mathbf{a}}^T(k)\mathbf{W} \qquad (6)$$

where $\mathbf{W}$ is a $(Np \times Np)$-dimensional transformation matrix which will be described in Section 3.

The constrained optimization problem can be solved using Lagrange multipliers. The Lagrangian function that we have to minimize is the following :

$$\Phi(k) = \textstyle\sum_{i=1}^{p} \mathbf{a}_i^T(k)\tilde{\mathbf{R}}_N\mathbf{a}_i(k) - \sum_{i=1}^{p}\sum_{j=1\ i\neq j}^{p} \lambda_{ij}\cdot$$
$$\sum_{l=0}^{N-1}[\tilde{a}(k)]_{p\cdot l+j,i} + \sum_{i=1}^{p}\lambda_{ii}(1 - \sum_{l=0}^{N-1}[\tilde{a}(k)]_{p\cdot l+i,i}) \qquad (7)$$

where $\mathbf{a}_i(k)$ is the i-th column of $\tilde{\mathbf{a}}(k)$ and λ_{ij} the Lagrange multipliers. Unlike the unconstrained L filter, the optimal location invariant filter coefficients do not depend on the noise-free signal but only on the correlation matrix $\tilde{\mathbf{R}}_N = E[\tilde{\mathbf{n}}^T\tilde{\mathbf{n}}]$ of the $(1 \times Np)$-dimensional vector $\tilde{\mathbf{n}}(k)$ of the ordered noise vectors inside the filter window :

$$\tilde{\mathbf{n}}(k) = [\mathbf{n}_{[1]}^T(k),\dots,\mathbf{n}_{[N]}^T(k)] = \tilde{\mathbf{x}}(k) - [\mathbf{s}^T,\dots,\mathbf{s}^T]. \qquad (8)$$

3 ADAPTIVE MULTICHANNEL L FILTERS

Using an LMS-like methodology we have devised coefficient adaptation formulas for both the unconstrained and the location invariant filters. The coefficients derived from these formulas are guaranteed to converge towards the optimal coefficients. Unfortunately the dependency of the optimal unconstrained filter coefficients on the noise-free signal renders their adaptive counterparts unsuitable for applications where strong nonstationarites are present (e.g. color image processing) since in this case the adaptation algorithm will face difficulties in tracking the ever changing optimal coefficients. This paper would therefore deal merely with the location invariant adaptive filters.

The LMS algorithm uses instantaneous estimates for the input data correlation matrix and the cross-correlation vector between the input data and the desired response to calculate the gradient vector involved in the steepest descent coefficient adaptation algorithm [3]. Modifications of this algorithm to incorporate linear constraints on the coefficients have also been proposed. The adaptation procedure for the location invariant multichannel L filter coefficients was derived by appropriately modifying the constrained LMS algorithm proposed in [4]. The basic feature of this algorithm is that during the evaluation of the coefficient matrix at the current iteration, the coefficient matrix evaluated at the previous iteration is not assumed to precisely satisfy the imposed constraint. This gives to the algorithm the ability to correct small deviations from the constraint that are caused by computational errors (roundoff, quantization, truncation errors) thus preventing them from accumulating during adaptation and causing instability.

The steepest descent algorithm for the adaptation of the location invariant filter coefficients can be expressed as follows :

$$\mathbf{a}_i(k+1) = \mathbf{a}_i(k) - \frac{1}{2}\mu\nabla_{\mathbf{a}_i}\Phi(k) \quad i = 1,\dots,p \qquad (9)$$

where $\nabla_{\mathbf{a}_i}\Phi(k)$ is the gradient of (7) with respect to $\mathbf{a}_i(k)$ and μ is the adaptation step. The gradient $\nabla_{\mathbf{a}_i}\Phi(k)$ was found to be :

$$\nabla_{\mathbf{a}_i}\Phi(k) = 2\tilde{\mathbf{R}}_N\mathbf{a}_i(k) - \tilde{\lambda}_i \qquad (10)$$

where

$$\tilde{\lambda}_i = \begin{bmatrix} \underline{\lambda}_i \\ \vdots \\ \underline{\lambda}_i \end{bmatrix} \qquad (11)$$

$$\underline{\lambda}_i = [\ \lambda_{i1} \ \dots \ \lambda_{ip}\ ]^T. \qquad (12)$$

Combining (9), (10) we get :

$$\tilde{\mathbf{a}}^T(k+1) = \tilde{\mathbf{a}}^T(k) - \mu\tilde{\mathbf{a}}^T(k)\tilde{\mathbf{R}}_N + \frac{1}{2}\mu\tilde{\lambda}^T \qquad (13)$$

$$\tilde{\lambda} = [\ \tilde{\lambda}_1 \ \dots \ \tilde{\lambda}_p\ ]. \qquad (14)$$

In order to proceed we have to evaluate $\tilde{\lambda}$. This can be done by forcing $\tilde{\mathbf{a}}(k+1)$ to comply with the constraints

(3) or equivalently forcing $\tilde{\mathbf{b}}(k+1)$ to satisfy (4). By postmultiplying (13) with $\mathbf{W}$ and using (6) we obtain the recursion formula in terms of $\tilde{\mathbf{b}}$:

$$\tilde{\mathbf{b}}(k+1) = \tilde{\mathbf{b}}(k) - \mu\tilde{\mathbf{b}}(k)\mathbf{P} + \frac{1}{2}\mu\tilde{\lambda}^T\mathbf{W}. \qquad (15)$$

The matrix $\mathbf{P}$ can be obtained from $\tilde{\mathbf{R}}_N$ through the following rearrangement of its elements :

$$[P]_{N\cdot(i-1)+j,N\cdot(k-1)+l} = [\tilde{R}_N]_{i+p\cdot(j-1),k+p\cdot(l-1)}$$
$$i,k = 1,\ldots,p \;\; j,l = 1,\ldots,N \qquad (16)$$

where $[P]_{i,j}$, $[\tilde{R}_N]_{i,j}$ denote the elements of $\mathbf{P}$, $\tilde{\mathbf{R}}_N$ respectively. This rearrangement is the result of consecutive interchanges between certain rows and columns of $\tilde{\mathbf{R}}_N$. For example in the case $p=2$, $N=3$ the rearrangement can be done by the following sequence of column interchanges $(2 \leftrightarrow 3)$, $(5 \leftrightarrow 4)$, $(3 \leftrightarrow 4)$ followed by the same sequence of row interchanges. Each column (row) interchange can be performed by postmultiplying (premultiplying) the matrix $\tilde{\mathbf{R}}_N$ by a $Np \times Np$ identity matrix that has been subject to the same interchange. If we define by $\mathbf{I}_{(a,b)}$ the identity matrix whose rows a, b has been interchanged the rearrangement (16) for $p=2$, $N=3$ can be expressed as follows :

$$\mathbf{P} = \mathbf{W}^T\tilde{\mathbf{R}}_N\mathbf{W}, \quad \mathbf{W} = \mathbf{I}_{(2,3)}\mathbf{I}_{(5,4)}\mathbf{I}_{(3,4)} \qquad (17)$$

Using (4) and properties that stem from the special structure of $\mathbf{W}, \tilde{\lambda}$ one can solve (15) for $\frac{1}{2}\mu\tilde{\lambda}^T\mathbf{W}$:

$$\frac{1}{2}\mu\tilde{\lambda}^T\mathbf{W} = \frac{1}{N}(\mathbf{T} - \tilde{\mathbf{b}}(k)\mathbf{F} + \mu\tilde{\mathbf{b}}(k)\mathbf{PF}) \qquad (18)$$

$$\mathbf{F} = \mathbf{T}^T\mathbf{T}. \qquad (19)$$

By combining (15),(18) we derive the steepest descent adaptation formula for the location invariant L filter coefficients :

$$\tilde{\mathbf{b}}(k+1) = \tilde{\mathbf{b}}(k) - \mu\tilde{\mathbf{b}}(k)\mathbf{P} + \frac{1}{N}(\mathbf{T} - \tilde{\mathbf{b}}(k)\mathbf{F} + \mu\tilde{\mathbf{b}}(k)\mathbf{PF}) \qquad (20)$$

An estimate $\hat{\mathbf{P}}(k)$ for $\mathbf{P}$ can be obtained by replacing $\tilde{\mathbf{R}}_N$ in (17) by the instantaneous estimate $\tilde{\mathbf{n}}^T(k)\tilde{\mathbf{n}}(k)$. By doing so we get the LMS adaptation formula for the coefficients of the location invariant multichannel L filters :

$$\tilde{\mathbf{b}}(k+1) = \tilde{\mathbf{b}}(k) - \mu\tilde{\mathbf{b}}(k)\mathbf{W}^T\tilde{\mathbf{n}}^T(k)\tilde{\mathbf{n}}(k)\mathbf{W} +$$
$$\tfrac{1}{N}(\mathbf{T} - \tilde{\mathbf{b}}(k)\mathbf{F} + \mu\tilde{\mathbf{b}}(k)\mathbf{W}^T\tilde{\mathbf{n}}^T(k)\tilde{\mathbf{n}}(k)\mathbf{WF}) \qquad (21)$$

4 CONVERGENCE ANALYSIS

The convergence in the mean for the proposed adaptive location invariant filter i.e., the convergence of the mean coefficient matrix $E[\tilde{\mathbf{a}}(k)]$ towards the optimal coefficient matrix $\tilde{\mathbf{a}}_{opt}$ or, equivalently, the convergence of $E[\tilde{\mathbf{b}}(k)]$ towards $\tilde{\mathbf{b}}_{opt}$ has been studied. Assumptions analogous to the so-called fundamental assumption [3] that has been widely used for the study of the classical LMS algorithm were adopted. These include the assumption that the ordered input vectors $\tilde{\mathbf{x}}(k)$ at time k are independent of the input vectors $\tilde{\mathbf{x}}(k-1),\ldots,\tilde{\mathbf{x}}(0)$ in the previous time instants and the assumption that the input vectors $\tilde{\mathbf{x}}(k)$ at time k are independent from the desired response $\mathbf{s}(k)$ in previous instants.

Let us define the $(p \times Np)$ weight error vector $\mathbf{v}(k)$ with respect to $\tilde{\mathbf{b}}(k)$

$$\mathbf{v}(k) = \tilde{\mathbf{b}}(k) - \tilde{\mathbf{b}}_{opt}. \qquad (22)$$

By subtracting $\tilde{\mathbf{b}}_{opt}$ from both sides of (21) we get :

$$\mathbf{v}(k+1) = \mathbf{v}(k) - \tilde{\mathbf{b}}(k)(\frac{1}{N}\mathbf{F} + \mu\hat{\mathbf{P}}(k)(\mathbf{I}_{Np} - \frac{1}{N}\mathbf{F})) + \frac{1}{N}\mathbf{T} \qquad (23)$$

After some manipulation and by taking the statistical expectation of both sides and utilizing the fundamental assumption we obtain the following expression :

$$E[\mathbf{v}(k+1)] = E[\mathbf{v}(k)](\mathbf{I}_{Np} - \mu\mathbf{P})(\mathbf{I}_{Np} - \tfrac{1}{N}\mathbf{F}) -$$
$$\tfrac{1}{N}\tilde{\mathbf{b}}_{opt}\mathbf{F} - \mu\tilde{\mathbf{b}}_{opt}\mathbf{P}(\mathbf{I}_{Np} - \tfrac{1}{N}\mathbf{F}) + \tfrac{1}{N}\mathbf{T} \qquad (24)$$

It can be shown that $\frac{1}{N}\tilde{\mathbf{b}}_{opt}\mathbf{F} - \mu\tilde{\mathbf{b}}_{opt}\mathbf{P}(\mathbf{I}_{Np} - \frac{1}{N}\mathbf{F}) + \frac{1}{N}\mathbf{T} = 0$ and therefore :

$$E[\mathbf{v}(k+1)] = E[\mathbf{v}(k)](\mathbf{I}_{Np} - \mu\mathbf{P})(\mathbf{I}_{Np} - \frac{\mathbf{F}}{N}) \qquad (25)$$

or equivalently :

$$E[\mathbf{v}_i(k+1)] = E[\mathbf{v}_i(k)](\mathbf{I}_{Np} - (\mu\mathbf{P} + \frac{1}{N}\mathbf{F} - \frac{\mu}{N}\mathbf{PF})) \qquad (26)$$

where $\mathbf{v}_i(k)$ is the i-th row of $\mathbf{v}(k)$. This equation shows that for a specific j all $[\tilde{b}]_{ij}$, $i = 1,\ldots,p$ exhibit the same transient behavior. Equation (26) is exactly analogous to the one that governs the convergence of the mean weight error vector in the standard LMS algorithm [3]. Therefore, we can state that the convergence in the mean for the location invariant L filter is governed by the eigenvalues l_i of the matrix $\mathbf{V} = \mu\mathbf{P} + \frac{1}{N}\mathbf{F} - \frac{\mu}{N}\mathbf{PF}$. In order that the mean coefficient matrix converges towards $\tilde{\mathbf{b}}_{opt}$, l_i should lie in the interval $(0,2)$. Therefore μ should be chosen so that the above-mentioned condition is satisfied. The overall time constant τ_a for the location invariant filter coefficients is bounded by :

$$\frac{-1}{\ln(1-l_{max})} < \tau_a < \frac{-1}{\ln(1-l_{min})} \qquad (27)$$

where l_{min}, l_{max} are the smallest and the largest eigenvalues of matrix $\mathbf{V}$ respectively.

5 EXPERIMENTAL RESULTS

The proposed adaptive location invariant filter was tested in the filtering of noisy color images. The color image "Pepper" of dimensions 256×256 has been used in the experiments. The above-mentioned image was corrupted by additive multivariate contaminated Gaussian noise of the form :

$$(1-\epsilon)\cdot\mathbf{N}(m_{1R},m_{1G},m_{1B},\sigma_{1R},\sigma_{1G},\sigma_{1B},r_{RG},r_{GB},r_{RB})+$$

$$\epsilon\cdot\mathbf{N}(m_{2R},m_{2G},m_{2B},\sigma_{2R},\sigma_{2G},\sigma_{2B},r'_{RG},r'_{GB},r'_{RB}) \quad (28)$$

Two different sets of noise parameters were used in our experiments. In both sets the two Gaussian populations had zero mean i.e., $m_{1R}=m_{1G}=m_{1B}=m_{2R}=m_{2G}=m_{2B}=0$ and the following correlation coefficients $r_{RG}=r_{GB}=r_{RB}=r'_{RG}=r'_{GB}=r'_{RB}=0.5$. The contamination factor was $\epsilon=0.2$ for both cases. The standard deviation for the two populations in the first experiment (#1 in Table 1) were chosen $\sigma_{1R}=\sigma_{1G}=\sigma_{1B}=20$, $\sigma_{2R}=\sigma_{2G}=\sigma_{2B}=40$, whereas in the second experiment (#2 in Table 1) we used $\sigma_{1R}=\sigma_{1G}=\sigma_{1B}=20$, $\sigma_{2R}=\sigma_{2G}=\sigma_{2B}=60$ in order to obtain an impulsive-like noise. The reference vector $\overline{\mathbf{a}}$ for the location invariant filter was chosen to be the output of a 5×5 marginal α-trimmed mean filter with an appropriately chosen α. The same vector was also used as an estimate of $\mathbf{s}$ in order to evaluate the noise vectors $\tilde{\mathbf{n}}(k)$ in (21) using (8). The adaptation step was chosen to be $\mu=5\cdot 10^{-6}$ which results in eigenvalues for the matrix $\mathbf{V}$ that lie within the interval $(0,2)$. The total coefficient absolute error $\varepsilon(k)$:

$$\varepsilon(k)=\sum_{i=1}^{Np}\sum_{j=1}^{p}|[\tilde{a}(k)]_{i,j}-[\tilde{a}_{opt}]_{i,j}| \quad (29)$$

where k is a running index corresponding to the raster-scan way of traversing the noisy image can be seen in Fig. 1 for the first set of noise parameters. The 3×3 adaptive multichannel location invariant L filter was compared to the following filters of the same window size : optimal location invariant multichannel L filter, arithmetic mean, marginal median, vector median, marginal α-trimmed mean with properly chosen α value and optimal single-channel location invariant L filter acting separately on each channel. The performance of the various filters was compared with respect to the Noise Reduction Index (Table 1). The proposed adaptive location invariant multichannel filter achieved the highest noise reduction among all the filters in both cases, surpassing even the optimal location invariant multichannel L filter. This can be explained by the fact that the correlation matrix that is required for the evaluation of the optimal filter coefficients has been numerically evaluated and thus are subject to a small error. For both the optimal and the adaptive location invariant filters a second entry (marked #2) exists in Table 1. These entries correspond to a combination of the multichannel L filter with a marginal α-trimmed mean filter that operates on edge areas.

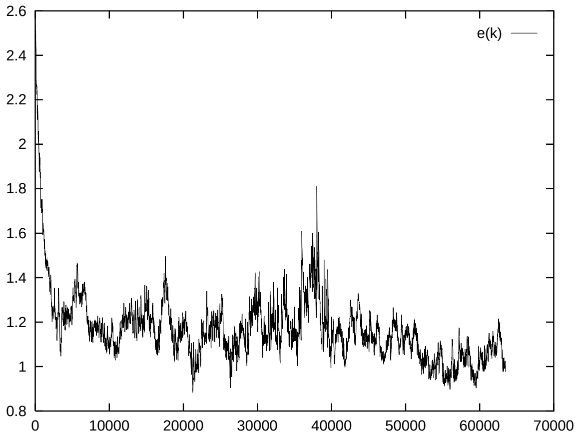

Figure 1: Total coefficient absolute error $\varepsilon(k)$ for the location invariant adaptive L filter applied in a raster-scan fashion on the noisy "Pepper" image.

Filter Type	Noise # 1	Noise #2
Adapt. Loc. Invariant L #1	-7.92	-9.79
Adapt. Loc. Invariant L #2	-8.20	-9.98
Optim. Loc. Invariant L #1	-7.85	-9.75
Optim. Loc. Invariant L #2	-8.14	-9.95
Arith. Mean	-7.28	-8.01
Marginal Median	-6.82	-8.57
L_1 Vect. Median	-5.29	-7.15
α-trimmed	-7.56	-9.15
single-channel L	-7.58	-9.17

Table 1: NRI obtained by various 3×3 filters in the filtering of color image "Pepper" corrupted by two types of contaminated Gaussian noise.

References

[1] N. Nikolaidis, I. Pitas, "Optimal Multichannel L Filters Based on Reduced Ordering", *1995 IEEE Workshop on Nonlinear Signal and Image Processing*, Neos Marmaras, Greece, pp. 507-10.

[2] N. Nikolaidis, I. Pitas, "Multichannel L Filters Based on Reduced Ordering", *IEEE Transactions on Circuits and Systems for Video Technology*, in press, 1996.

[3] S. Haykin, *Adaptive Filter Theory*, Second edition, Prentice Hall, Englewood Cliffs, NJ, 1991.

[4] O. L. Frost III, " An algorithm for linearly constrained adaptive array processing", *Proceedings of the IEEE*, vol 60, no 8, pp. 926-935.