Multichannel L Filters Based on Marginal Data Ordering

Constantine Kotropoulos, *Member, IEEE*, and Ioannis Pitas, *Senior Member, IEEE*

***Abstract*— The extension of single-channel nonlinear filters whose output is a linear combination of the order statistics of the input samples to the multichannel case is presented in this paper. The subordering principle of marginal ordering (M-ordering) is used for multivariate data ordering. Assuming a multichannel signal corrupted by additive white multivariate noise whose components are generally correlated, the coefficients of the multichannel L filter based on marginal ordering are chosen to minimize the output mean-squared-error (MSE) either subject to the constraints of unbiased or location-invariant estimation or without imposing any constraint. Both the case of a constant multichannel signal corrupted by additive white multivariate noise as well as the case of a nonconstant signal is considered. In order to test the performance of the designed multichannel marginal L filters, long-tailed multivariate distributions are required. The derivation and design of such a distribution, namely, the Laplacian (biexponential) distribution that belongs to Morgenstern's family in the 2-D case is discussed. It is shown by simulations that the proposed multichannel L filters perform better than other multichannel nonlinear filters such as the vector median, the marginal α-trimmed mean, the marginal median, the multichannel modified trimmed mean, the multichannel double-window trimmed mean, and the multivariate ranked-order estimator $\mathcal{R}_E$ proposed elsewhere as well as their single-channel counterparts.**

I. Introduction

MULTICHANNEL 1-D and 2-D signals appear frequently in practice, for example, in the cases involving multiple sources and receivers, as in geophysics, underwater acoustics, multiple-antenna transmission systems, and in the processing of color images and sequences of images. A multichannel signal is defined as a vector of components called channels, which are generally correlated and characterized by their joint probability density function (pdf). If each signal component is processed separately, this correlation is not utilized. Although transformation techniques such as the Karhunen–Loeve transformation can be used first to decorrelate the signal components in order to apply single-channel signal processing techniques afterwards; a more natural way is to apply multichannel signal processing techniques.

A class of nonlinear filters that has found extensive applications in digital signal and image processing are the L filters (sometimes also called order statistic filters) whose output is defined as a linear combination of the order statistics of the input sequence [1], [2]. The design of an optimal L filter for estimating a constant signal corrupted by additive white noise has originally been proposed in [1] and has been extended to the restoration of a Markov signal in additive white noise [3], [4] as well as in the case of dependent noise and arbitrary waveforms [5].

Recently, increasing attention has been given to nonlinear processing of vector-valued signals. Vector median operations have been derived from multidimensional exponential pdfs using the maximum likelihood estimate approach in [6]. The use of ordering of multivariate data in multichannel signal processing has been described in [7]. Various multichannel estimators such as the marginal median, the marginal α-trimmed mean, and modified trimmed mean filters have been proposed in [7] and [8]. The ordering of multivariate data according to their respective distance from the population centroid has led to ranked-order-type estimators for multivariate image fields in [9]. Another class of multichannel L filters based on radial medians has been proposed in [10].

The main contribution of the paper is the design of multichannel L filters based on marginal ordering (M ordering) using the mean-squared-error (MSE) as fidelity criterion. M ordering implies independent data ordering in each channel. We assume that a multichannel signal is corrupted by additive white multivariate noise, which generally exhibits correlation between different channels. The unconstrained minimization of the MSE is treated first. Structural constraints such as unbiasedness and location invariance are also incorporated in the minimization procedure. The unconstrained minimization is shown that it leads to a global minimum. The design procedure involves moments of the order statistics of input samples derived from the same channel as well as from different channels. The theoretical framework required for the computation of the above-mentioned moments is outlined, and a discrete algorithm for their computation is derived based on vector quantization. In order to test the performance of the designed multichannel marginal L filters, long-tailed multivariate distributions are required. The derivation and design of such a distribution, namely, the Laplacian (biexponential) distribution that belongs to Morgenstern's family in the two-channel case is discussed. The noise-reduction capability of the designed multichannel nonlinear filters is examined for bivariate distributions ranging from the short-tailed to the long-tailed ones. The following bivariate pdfs are considered: uniform, joint Gaussian, contaminated Gaussian, and Laplacian distributions. It is shown by simulations that the proposed

Manuscript received December 10, 1991; revised December 16, 1993. This work was supported by a scholarship from the State Scholarship Foundation of Greece and the Bodosaki Foundation. The associate editor coordinating the review of this paper and approving it for publication was Prof. Gonzalo Arce.

The authors are with the Department of Electrical and Computer Engineering, University of Thessaloniki, Thessaloniki, Greece.

IEEE Log Number 9403722.

1053-587X/94$04.00

multichannel L filters perform better than other multichannel nonlinear filters such as the vector median [6], the marginal α-trimmed mean [7], [8], the marginal median [7], [8], the multichannel modified trimmed mean [8] and the multichannel double-window trimmed mean [8] the multivariate ranked-order estimator $\mathcal{R}_E$ [9], and their single-channel counterparts [1].

The work presented in this paper extends previously reported work [6]–[8]. We have tried to make the paper sufficiently self-contained to be read without extensive use of references. The outline of the paper is as follows. After a short revision of the basic concepts of multivariate data ordering, we proceed to the design of multichannel marginal L filters in Section II. The calculation of the moments of the multivariate order statistics is examined in Section III. The derivation of the bivariate Laplacian distribution is treated in Section IV. Simulation examples are included in Section V. Conclusions are drawn in Section VI.

II. Multichannel L Filters Based on M Ordering

Univariate ordering operations have proven to be useful in robust estimation techniques because if outliers exist in a set of input samples, they will generally be located in the extreme sorted data [11], [12]. Therefore, the outliers can be isolated by sorting the input data and can be deemphasized or discarded before the desired estimate is computed by weighting the order statistics appropriately.

The notion of data ordering cannot be extended in a straightforward manner from the univariate case to the multivariate one. There are several ways to order multivariate data, but none of them is unambiguous nor universally accepted. Specifically, the following four so-called subordering principles are discussed in [14]: marginal ordering (M ordering), partial ordering, conditional ordering, and reduced ordering. In the sequel, we shall confine ourselves to the definition of M ordering. For more details, the interested reader can refer to the statistical literature [14], [15] or [7]–[9].

Let $\mathbf{x}_1, \ldots, \mathbf{x}_N$ be a random sample of N observations of a p-dimensional random variable $\mathbf{X}$. Each vector-valued observation $\mathbf{x}_i = (x_{i1}, x_{i2}, \ldots, x_{ip})^T$ belongs to a p-dimensional space denoted by $\mathcal{R}^p$.

Definition 1: The M-ordering scheme orders each of the vector components independently, yielding

$$x_{j(1)} \leq x_{j(2)} \leq \cdots \leq x_{j(N)} \quad j = 1, \ldots, p \tag{1}$$

i.e., the vector-valued observations are ordered along each of the p dimensions (or channels) independently.

$x_{1(1)}, x_{2(1)}, \ldots, x_{p(1)}$ are the minimal elements in each channel. $x_{1(N)}, x_{2(N)}, \ldots, x_{p(N)}$ are the maximal elements in each channel. By definition, the $(i_1, i_2, \ldots, i_p)$-marginal order statistic is the following $(p \times 1)$ vector: $\mathbf{x}_{(i_1, i_2, \ldots, i_p)} = (x_{1(i_1)}, x_{2(i_2)}, \ldots, x_{p(i_p)})^T$ where $1 \leq i_j \leq N$; $j = 1, \ldots, p$. The probability distributions of the p-dimensional marginal order statistics have been derived in [7] and [8]. Having described the M-ordering principle, we proceed to the definition and design of multichannel L filters based on the above-mentioned subordering principle.

Definition 2: The output of a p-channel marginal L filter of length N operating on a sequence of p-dimensional vectors $\{\mathbf{x}(k)\}$ for N odd is given by

$$\mathbf{y}(k) \stackrel{\text{def}}{=} \mathbf{T}[\mathbf{x}(k)] = \sum_{j=1}^{p} \mathbf{A}_j \tilde{\mathbf{x}}_j(k) \tag{2}$$

where $\mathbf{A}_j$ are appropriate $(p \times N)$ coefficient matrices, and $\tilde{\mathbf{x}}_j(k) = (x_{j(1)}(k), \ldots, x_{j(N)}(k))^T$ are the $(N \times 1)$ vectors of the order statistics along each channel.

It can easily be proven that the Definition 2 is fully equivalent to the alternative definition given in [7].

Let us suppose that the observed p-dimensional signal $\{\mathbf{x}(k)\}$ can be expressed as the sum of a known p-dimensional signal $\mathbf{s}(k)$ and a noise vector sequence $\{\mathbf{n}(k)\}$ of zero-mean vector having the same dimensionality, i.e., $\mathbf{x}(k) = \mathbf{s}(k) + \mathbf{n}(k)$. The noise vector $\mathbf{n}(k) = (n_1(k), \ldots, n_p(k))^T$ is a p-dimensional vector of random variables characterized by the joint pdf of its components, which are assumed to be correlated in the general case. In addition, we assume that the noise vectors at different time instants are independent identically distributed (i.i.d.) and that at every time instant, the signal $\mathbf{s}(k)$ and the noise vector $\mathbf{n}(k)$ are uncorrelated.

We shall design the p-channel marginal L filter that operates on the p-dimensional observed signal $\{\mathbf{x}(k)\}$ and is the optimal estimator of $\mathbf{s}(k)$ by using the MSE between $\mathbf{s}(k)$ and the output of the p-channel marginal L filter as fidelity criterion. Strictly speaking, we define as MSE the trace of the mean-squared-error matrix. The case $p = 2$ is treated first for notation simplicity, and generalizations are deduced in the sequel. For $p = 2$, the MSE $\varepsilon = E[(\mathbf{y}(k) - \mathbf{s}(k))^T (\mathbf{y}(k) - \mathbf{s}(k))]$ is given by

$$\begin{aligned} \varepsilon = \mathrm{E}\left[\sum_{i=1}^{2}\sum_{j=1}^{2} \tilde{\mathbf{x}}_i^T(k) \mathbf{A}_i^T \mathbf{A}_j \tilde{\mathbf{x}}_j(k)\right] \\ - 2\mathbf{s}^T(k)\mathrm{E}\left[\sum_{i=1}^{2} \mathbf{A}_i \tilde{\mathbf{x}}_i(k)\right] + \mathbf{s}^T(k)\mathbf{s}(k). \end{aligned} \tag{3}$$

In the following, the time index k will be suppressed without any lack of generality. The $(2 \times N)$ matrices $\mathbf{A}_i$ can be partitioned as follows:

$$\mathbf{A}_i = \begin{bmatrix} \mathbf{a}_{i1}^T \\ \mathbf{a}_{i2}^T \end{bmatrix} \tag{4}$$

where $\mathbf{a}_{il}^T, l = 1, 2$ are $(1 \times N)$ row vectors corresponding to the rows of matrix $\mathbf{A}_i$. Let $\mathbf{R}_{ji}$ denote the correlation matrix of the ordered input samples in channels j and i, i.e., $\mathbf{R}_{ji} = \mathrm{E}[\tilde{\mathbf{x}}_j \tilde{\mathbf{x}}_i^T]$. Since $\mathbf{R}_{ii}, i = 1, 2$ consist of moments of the order statistics from a univariate population, they will be called autocorrelation matrices of the order statistics. The remaining matrices $\mathbf{R}_{12}$ and $\mathbf{R}_{21}$ are related by transposition, i.e., $\mathbf{R}_{12} = \mathbf{R}_{21}^T$. Therefore, we are dealing only with $\mathbf{R}_{12}$. Its elements are product moments of the order statistics from a bivariate population. In addition, let $\underline{\mu}_j, j = 1, 2$ denote the mean vector of the order statistics in channel j, i.e.

$$\underline{\mu}_j \stackrel{\text{def}}{=} \left(\mathrm{E}[x_{j(1)}], \mathrm{E}[x_{j(2)}], \ldots, \mathrm{E}[x_{j(N)}]\right)^T. \tag{5}$$

By using the partitioning (4) and the definition of $\mathbf{R}_{ji}$, the first term of the right-hand side of (3) is rewritten as

$$\begin{aligned}\mathrm{E}\left[\sum_{i=1}^{2}\sum_{j=1}^{2}\tilde{\mathbf{x}}_i^T(k)\mathbf{A}_i^T\mathbf{A}_j\tilde{\mathbf{x}}_j(k)\right] &= \sum_{i=1}^{2}\sum_{j=1}^{2}\mathrm{tr}\left[\mathbf{A}_j\mathbf{R}_{ji}\mathbf{A}_i^T\right]\\ &= \sum_{i=1}^{2}\sum_{j=1}^{2}\left(\mathbf{a}_{j1}^T\mathbf{R}_{ji}\mathbf{a}_{i1}+\mathbf{a}_{j2}^T\mathbf{R}_{ji}\mathbf{a}_{i2}\right)\end{aligned} \tag{6}$$

where tr[ ] stands for the trace of the bracketed matrix. After some algebraic manipulation, (6) can be rewritten in the form:

$$\begin{aligned}\mathrm{E}\left[\sum_{i=1}^{2}\sum_{j=1}^{2}\tilde{\mathbf{x}}_i^T(k)\mathbf{A}_i^T\mathbf{A}_j\tilde{\mathbf{x}}_j(k)\right] &= [\mathbf{a}_{11}^T \mid \mathbf{a}_{21}^T]\hat{\mathbf{R}}_2\begin{bmatrix}\mathbf{a}_{11}\\ \mathbf{a}_{21}\end{bmatrix}\\ &\quad + [\mathbf{a}_{12}^T \mid \mathbf{a}_{22}^T]\hat{\mathbf{R}}_2\begin{bmatrix}\mathbf{a}_{12}\\ \mathbf{a}_{22}\end{bmatrix}\end{aligned} \tag{7}$$

where $\hat{\mathbf{R}}_2$ is the following matrix:

$$\hat{\mathbf{R}}_2 = \begin{bmatrix}\mathbf{R}_{11} & \mathbf{R}_{12}\\ \mathbf{R}_{12}^T & \mathbf{R}_{22}\end{bmatrix}. \tag{8}$$

By substituting (4)–(8) into (3), we obtain

$$\begin{aligned}\varepsilon &= [\mathbf{a}_{11}^T \mid \mathbf{a}_{12}]\hat{\mathbf{R}}_2\begin{bmatrix}\mathbf{a}_{11}\\ \mathbf{a}_{21}\end{bmatrix} + [\mathbf{a}_{12}^T \mid \mathbf{a}_{22}^T]\hat{\mathbf{R}}_2\begin{bmatrix}\mathbf{a}_{12}\\ \mathbf{a}_{22}\end{bmatrix}\\ &\quad - 2\mathbf{s}^T\begin{bmatrix}\mathbf{a}_{11}^T\\ \mathbf{a}_{12}^T\end{bmatrix}\underline{\mu}_1 - 2\mathbf{s}^T\begin{bmatrix}\mathbf{a}_{21}^T\\ \mathbf{a}_{22}^T\end{bmatrix}\underline{\mu}_2 + \mathbf{s}^T\mathbf{s}.\end{aligned} \tag{9}$$

In the general p-dimensional case, the MSE is given by

$$\varepsilon = \left\{\sum_{i=1}^{p}\mathbf{a}_{(i)}^T\hat{\mathbf{R}}_p\mathbf{a}_{(i)}\right\} - 2\mathbf{s}^T\begin{bmatrix}\mathbf{a}_{(1)}^T\\ \mathbf{a}_{(2)}^T\\ \vdots\\ \mathbf{a}_{(p)}^T\end{bmatrix}\hat{\underline{\mu}}_p + \mathbf{s}^T\mathbf{s} \tag{10}$$

where $\mathbf{a}_{(i)} = (\mathbf{a}_{1i}^T \quad \mathbf{a}_{2i}^T \quad \cdots \quad \mathbf{a}_{pi}^T)^T$, $\hat{\underline{\mu}}_p = (\underline{\mu}_1^T\ \underline{\mu}_2^T \cdots \underline{\mu}_p^T)^T$ and

$$\hat{\mathbf{R}}_p = \begin{bmatrix}\mathbf{R}_{11} & \mathbf{R}_{12} & \cdots & \mathbf{R}_{1p}\\ \mathbf{R}_{12}^T & \mathbf{R}_{22} & \cdots & \mathbf{R}_{2p}\\ \vdots & & \ddots & \vdots\\ \mathbf{R}_{1p}^T & \mathbf{R}_{2p}^T & \cdots & \mathbf{R}_{pp}\end{bmatrix}. \tag{11}$$

In the sequel, we shall treat first the unconstrained minimization of the MSE, and then, we shall impose constraints on the output of the multichannel marginal L filter.

A. Unconstrained Solution

Let us consider the case $p = 2$. Minimizing (9) over $\mathbf{a}_{ji}$, $i = 1, 2$ is a quadratic optimization problem that has a unique solution, provided that the symmetric matrix $\hat{\mathbf{R}}_2$ is positive definite. The components of the vectors of the order statistics along each channel $\tilde{\mathbf{x}}_j$, $j = 1, 2$ are linearly independent variables with probability 1 (see pp. 179–180 of [16]) due to the independence of the observations at different time instants. Thus, the diagonal submatrices of $\hat{\mathbf{R}}_2$, $\mathbf{R}_{jj}$, $j = 1, 2$, are positive definite with probability 1. The $2N$ random variables that form the vector of the order statistics from both channels $(\tilde{\mathbf{x}}_1^T\tilde{\mathbf{x}}_2^T)^T$ are linearly dependent in the general case, due to the correlation that exists between the ordered samples that correspond to the same time instant. Consequently, $\hat{\mathbf{R}}_2$ is positive semi-definite in general. We shall assume that $\hat{\mathbf{R}}_2$ is not singular, i.e., that $\hat{\mathbf{R}}_2$ is indeed positive definite. Such an assumption has been verified in all simulations performed in Section V.

Equating the derivatives of ε with respect to $\mathbf{a}_{ji}$ with zero, i.e., $\frac{\partial\varepsilon}{\partial\mathbf{a}_{ji}} = \mathbf{0}$, the following two sets of equations result:

$$\begin{aligned}\hat{\mathbf{R}}_2\mathbf{a}_{(1)}^* &= s_1\hat{\underline{\mu}}_2\\ \hat{\mathbf{R}}_2\mathbf{a}_{(2)}^* &= s_2\hat{\underline{\mu}}_2\end{aligned} \tag{12}$$

which yield the optimal two-channel marginal L filter coefficients, provided that all inverse matrices exist [17], [18]:

$$\begin{aligned}\mathbf{a}_{(1)}^* &= s_1\hat{\mathbf{R}}_2^{-1}\hat{\underline{\mu}}_2 = s_1\begin{bmatrix}\mathbf{R}_{11}^{-1}+\mathbf{F}\mathbf{E}^{-1}\mathbf{F}^T & -\mathbf{F}\mathbf{E}^{-1}\\ -\mathbf{E}^{-1}\mathbf{F}^T & \mathbf{E}^{-1}\end{bmatrix}\hat{\underline{\mu}}_2\\ \mathbf{a}_{(2)}^* &= \frac{s_2}{s_1}\mathbf{a}_{(1)}^*\end{aligned} \tag{13}$$

where $\mathbf{E} = \mathbf{R}_{22} - \mathbf{R}_{12}^T\mathbf{R}_{11}^{-1}\mathbf{R}_{12}$ and $\mathbf{F} = \mathbf{R}_{11}^{-1}\mathbf{R}_{12}$. The resulting minimum MSE (MMSE) is given by $\varepsilon_{\min} = (1-\Delta)\mathbf{s}^T\mathbf{s}$, where $\Delta = \hat{\underline{\mu}}_2^T\hat{\mathbf{R}}_2^{-1}\hat{\underline{\mu}}_2$. The fact that ε is always nonnegative implies that $\varepsilon_{\min} \geq 0$. Therefore, $0 \leq \Delta \leq 1$.

In the general p-dimensional case, the minimization of the MSE (10) results in the following p sets of equations:

$$\hat{\mathbf{R}}_p\mathbf{a}_{(m)}^* = s_m\hat{\underline{\mu}}_p \qquad m = 1, \ldots, p \tag{14}$$

which yield the optimal p-channel marginal L filter coefficients:

$$\begin{aligned}\mathbf{a}_{(1)}^* &= s_1\hat{\mathbf{R}}_p^{-1}\hat{\underline{\mu}}_p\\ \mathbf{a}_{(m)}^* &= \frac{s_m}{s_1}\mathbf{a}_{(1)}^* \quad m = 2, \ldots, p.\end{aligned} \tag{15}$$

The MMSE associated with the optimal coefficients (15) is $\varepsilon_{\min} = (1 - \hat{\underline{\mu}}_p^T\hat{\mathbf{R}}_p^{-1}\hat{\underline{\mu}}_p)\mathbf{s}^T\mathbf{s}$. In (13) or (15), the optimal coefficients for the unconstrained multichannel marginal L filter depend on the knowledge of the signal $\mathbf{s}$ (to be estimated). In addition, the distributional model (i.e., the marginal and joint probability/cumulative density functions) of the components of the input vector-valued signal $\mathbf{x}(k)$ must be known in order to calculate $\hat{\mathbf{R}}_p$ and $\hat{\underline{\mu}}_p$, as is analyzed in Sections III-B and C. In many practical applications, the signal s to be estimated is unknown, unless the detection of a known signal in noise is investigated. Furthermore, in general, the distributional model of input vector data is unknown. Even in the case that the joint probability density function of noise samples $f_{n_1,\ldots,n_p}(n_1,\ldots,n_p)$ is known, the joint probability density function of the input vector-valued observations depends explicitly on $\mathbf{s}$ since $f_{x_1,\ldots,x_p}(x_1,\ldots,x_p) = f_{n_1,\ldots,n_p}(x_1 - s_1,\ldots,x_p - s_p)$. Efficient procedures for estimating $\mathbf{s}$ and for the calculation of $\hat{\mathbf{R}}_p$ and $\hat{\underline{\mu}}_p$ based on estimates of the marginal and joint probability density function of the input vector data are developed in Section II-D.

B. Unbiased Solution

In the univariate case, L filters are designed by imposing local structural constraints on the output of the L filter. Two types of constraints have been incorporated in the design of single-channel L filters [1], [4], [5]: unbiasedness and location-invariance. The unbiasedness constraint is examined in this section and the location-invariance will be considered in Section 2-C.

Definition 3: A multichannel marginal L filter is said to be an unbiased multichannel estimator of location, if $\mathrm{E}[\mathbf{y}(k)] = \mathbf{s}$, that is, the coefficients of the multichannel marginal L filter should satisfy the set of equations

$$\mathbf{a}_{(i)}^T\hat{\underline{\mu}}_p = s_i \qquad i = 1,\ldots,p. \tag{16}$$

The case $p = 2$ is considered first. Under the set of constraints (16), the MSE given by (9) is rewritten as

$$\varepsilon_{\mathrm{unb}} = \mathbf{a}_{(1)}^T\hat{\mathbf{R}}_2\mathbf{a}_{(1)} + \mathbf{a}_{(2)}^T\hat{\mathbf{R}}_2\mathbf{a}_{(2)} - \mathbf{s}^T\mathbf{s}. \tag{17}$$

The minimization of (17) subject to (16) can be solved by using Lagrange multipliers. The Lagrangian function is given by

$$\begin{aligned}\Lambda(\mathbf{a}_{ji},\lambda_j;j,i = 1,2) &= \mathbf{a}_{(1)}^T\hat{\mathbf{R}}_2\mathbf{a}_{(1)} + \mathbf{a}_{(2)}^T\hat{\mathbf{R}}_2\mathbf{a}_{(2)} - \mathbf{s}^T\mathbf{s}\\ &\quad + \lambda_1\left(s_1 - \mathbf{a}_{(1)}^T\hat{\underline{\mu}}_2\right) + \lambda_2\left(s_2 - \mathbf{a}_{(2)}^T\hat{\underline{\mu}}_2\right).\end{aligned} \tag{18}$$

Differentiating $\Lambda(\mathbf{a}_{ji},\lambda_j;j,i = 1,2)$ with respect to $\mathbf{a}_{ji}, j,i = 1,2$ and equating the result with zero yields the following two sets of equations in terms of the Lagrange multipliers λ_1,λ_2:

$$\begin{aligned}\hat{\mathbf{R}}_2\mathbf{a}_{(1)}^* &= \frac{\lambda_1}{2}\hat{\underline{\mu}}_2\\ \hat{\mathbf{R}}_2\mathbf{a}_{(2)}^* &= \frac{\lambda_2}{2}\hat{\underline{\mu}}_2.\end{aligned} \tag{19}$$

The Lagrange multipliers λ_1,λ_2 can be found by solving (19) with respect to $\mathbf{a}_{(1)}$ and $\mathbf{a}_{(2)}$ and by substituting these vectors into (16). Provided that $\mathbf{R}_{12}$ and $\mathbf{E}$ are nonsingular, the Lagrange multipliers are given by $\lambda_1 = \frac{2s_1}{\Delta}$ and $\lambda_2 = \frac{s_2}{s_1}\lambda_1$. Then, the optimal coefficients of the unbiased two-channel marginal L filter can be expressed as

$$\begin{aligned}\mathbf{a}_{(1)}^* &= \frac{s_1}{\Delta}\hat{\mathbf{R}}_2^{-1}\hat{\underline{\mu}}_2\\ \mathbf{a}_{(2)}^* &= \frac{s_2}{s_1}\mathbf{a}_{(1)}^*.\end{aligned} \tag{20}$$

The MMSE associated with the optimal coefficients (20) is $(\varepsilon_{\mathrm{unb}})_{\min} = \frac{1-\Delta}{\Delta}\mathbf{s}^T\mathbf{s}$. If $\hat{\mathbf{R}}_2$ is positive definite, then $\Delta > 0$. It has also been shown in Section II-A that $\Delta \le 1$. Therefore, the MMSE associated with the optimal unbiased two-channel marginal L filter is always greater than the MMSE produced by the optimal unconstrained two-channel marginal L filter.

It can be easily shown that the optimal coefficients of the unbiased p-channel marginal L filter are given by

$$\begin{aligned}\mathbf{a}_{(1)}^* &= \frac{s_1}{\hat{\underline{\mu}}_p^T\hat{\mathbf{R}}_p^{-1}\hat{\underline{\mu}}_p}\hat{\mathbf{R}}_p^{-1}\hat{\underline{\mu}}_p\\ \mathbf{a}_{(m)}^* &= \frac{s_m}{s_1}\mathbf{a}_{(1)}^* \qquad m = 2,\ldots,p.\end{aligned} \tag{21}$$

and the MMSE associated with the optimal coefficients (21) is

$$(\varepsilon_{\mathrm{unb}})_{\min} = \frac{1 - \hat{\underline{\mu}}_p^T\hat{\mathbf{R}}_p^{-1}\hat{\underline{\mu}}_p}{\hat{\underline{\mu}}_p^T\hat{\mathbf{R}}_p^{-1}\hat{\underline{\mu}}_p}\mathbf{s}^T\mathbf{s}. \tag{22}$$

The need for an estimate $\hat{\mathbf{s}}(k)$ of $\mathbf{s}$ as well as for the design of unbiased multichannel L filters based on estimates of the marginal and joint probability density functions of input vector-valued observations is recognized in this case as well.

C. Location-Invariant Solution

Definition 4: A multichannel marginal L filter is said to be location invariant if its output is able to track small perturbations of its input, i.e., if $\mathbf{x}'(k) = \mathbf{x}(k) + \mathbf{b}$, then

$$\mathbf{y}'(k) = \mathbf{T}[\mathbf{x}'(k)] = \mathbf{y}(k) + \mathbf{b} \tag{23}$$

where $\mathbf{y}(k) = \mathbf{T}[\mathbf{x}(k)]$.

Such an L filter is also called smoothing L filter because it preserves zero frequency or dc signals [5]. The definition of a location-invariant multichannel marginal L filter (23) yields the following set of constraints imposed on the filter coefficients:

$$\begin{aligned}\mathbf{e}^T\mathbf{a}_{jj} &= 1 \quad \forall j, \quad j = 1,\ldots,p\\ \mathbf{e}^T\mathbf{a}_{ji} &= 0 \quad \forall i, \quad i \ne j, \quad j = 1,\ldots,p\end{aligned} \tag{24}$$

where $\mathbf{e}$ denotes the $(N \times 1)$ unitary vector, i.e., $\mathbf{e} = (1,1,\ldots,1)^T$. By incorporating (24) into (9), we obtain $\varepsilon_{\mathrm{loc}} = \mathbf{a}_{(1)}^T\check{\mathbf{R}}_2\mathbf{a}_{(1)} + \mathbf{a}_{(2)}^T\check{\mathbf{R}}_2\mathbf{a}_{(2)}$, where

$$\check{\mathbf{R}}_2 = \begin{bmatrix}\tilde{\mathbf{R}}_{11} & \tilde{\mathbf{R}}_{12}\\ \tilde{\mathbf{R}}_{12}^T & \tilde{\mathbf{R}}_{22}\end{bmatrix} \tag{25}$$

with $\tilde{\mathbf{R}}_{ji} = \mathrm{E}[\tilde{\mathbf{n}}_j\tilde{\mathbf{n}}_i^T]$, $j,i = 1,2$, and $\tilde{\mathbf{n}}_l = (n_{l(1)},\ldots,n_{l(N)})^T$, $l = 1,2$. The minimization of $\varepsilon_{\mathrm{loc}}$ subject to (24) is formulated as minimization of the following Lagrangian function:

$$\begin{aligned}\Lambda(\mathbf{a}_{ji},\lambda_{ji};j,i = 1,2) &= \mathbf{a}_{(1)}^T\check{\mathbf{R}}_2\mathbf{a}_{(1)} + \mathbf{a}_{(2)}^T\check{\mathbf{R}}_2\mathbf{a}_{(2)}\\ &\quad + \sum_{j=1}^{2}\left\{\lambda_{jj}(1 - \mathbf{e}^T\mathbf{a}_{jj}) - \sum_{i=1 i\ne j}^{2}\lambda_{ji}\mathbf{e}^T\mathbf{a}_{ji}\right\}\end{aligned} \tag{26}$$

where $\lambda_{ji}, j,i = 1$, and 2 are Lagrange multipliers. Differentiating $\Lambda(\mathbf{a}_{ji},\lambda_{ji};j,i = 1,2)$ with respect to $\mathbf{a}_{ji}$ and equating the result with zero yields the following two sets of equations in terms of λ_{ji}:

$$\check{\mathbf{R}}_2\mathbf{a}^*_{(1)} = \frac{1}{2}\begin{bmatrix}\lambda_{11}\mathbf{e}\\ \lambda_{21}\mathbf{e}\end{bmatrix}$$
$$\check{\mathbf{R}}_2\mathbf{a}^*_{(2)} = \frac{1}{2}\begin{bmatrix}\lambda_{12}\mathbf{e}\\ \lambda_{22}\mathbf{e}\end{bmatrix}. \tag{27}$$

The Lagrange multipliers can be found by solving (27) with respect to $\mathbf{a}^*_{(1)}$ and $\mathbf{a}^*_{(2)}$ and substituting the results into (24). Let us define [17], [18]

$$\begin{aligned}\mathbf{P}_{11} &= (\tilde{\mathbf{R}}_{11} - \tilde{\mathbf{R}}_{12}\tilde{\mathbf{R}}_{22}^{-1}\tilde{\mathbf{R}}_{12}^T)^{-1},\\ \mathbf{P}_{12} &= (\tilde{\mathbf{R}}_{12}^T - \tilde{\mathbf{R}}_{22}\tilde{\mathbf{R}}_{12}^{-1}\tilde{\mathbf{R}}_{11})^{-1}\\ \mathbf{P}_{21} &= \mathbf{P}_{12}^T,\\ \mathbf{P}_{22} &= (\tilde{\mathbf{R}}_{22} - \tilde{\mathbf{R}}_{12}^T\tilde{\mathbf{R}}_{11}^{-1}\tilde{\mathbf{R}}_{12})^{-1}\end{aligned} \tag{28}$$

and

$$\mathcal{D} = \det\begin{pmatrix}\mathbf{e}^T\mathbf{P}_{11}\mathbf{e} & \mathbf{e}^T\mathbf{P}_{12}\mathbf{e}\\ \mathbf{e}^T\mathbf{P}_{21}\mathbf{e} & \mathbf{e}^T\mathbf{P}_{22}\mathbf{e}\end{pmatrix} \tag{29}$$

where det() stands for the determinant of the matrix inside parentheses. If $\tilde{\mathbf{R}}_{12}$ and $(\tilde{\mathbf{R}}_{22} - \tilde{\mathbf{R}}_{12}^T\tilde{\mathbf{R}}_{11}^{-1}\tilde{\mathbf{R}}_{12})$ are nonsingular, the Lagrange multipliers $\lambda_{ji} j, i = 1,$, and 2 are given by

$$\begin{aligned}\lambda_{11} &= 2\frac{\mathbf{e}^T\mathbf{P}_{22}\mathbf{e}}{\mathcal{D}}, & \lambda_{21} &= -2\frac{\mathbf{e}^T\mathbf{P}_{21}\mathbf{e}}{\mathcal{D}}\\ \lambda_{12} &= -2\frac{\mathbf{e}^T\mathbf{P}_{12}\mathbf{e}}{\mathcal{D}}, & \lambda_{22} &= 2\frac{\mathbf{e}^T\mathbf{P}_{11}\mathbf{e}}{\mathcal{D}}.\end{aligned} \tag{30}$$

By using (27)–(30), the optimal coefficients of the location-invariant two-channel marginal L filter are

$$\begin{aligned}\mathbf{a}_{11} &= \left(\frac{\mathbf{e}^T\mathbf{P}_{22}\mathbf{e}}{\mathcal{D}}\right)\mathbf{P}_{11}\mathbf{e} - \left(\frac{\mathbf{e}^T\mathbf{P}_{21}\mathbf{e}}{\mathcal{D}}\right)\mathbf{P}_{12}\mathbf{e},\\ \mathbf{a}_{21} &= \left(\frac{\mathbf{e}^T\mathbf{P}_{21}\mathbf{e}}{\mathcal{D}}\right)\mathbf{P}_{21}\mathbf{e} - \left(\frac{\mathbf{e}^T\mathbf{P}_{21}\mathbf{e}}{\mathcal{D}}\right)\mathbf{P}_{22}\mathbf{e}\\ \mathbf{a}_{12} &= \left(\frac{\mathbf{e}^T\mathbf{P}_{12}\mathbf{e}}{\mathcal{D}}\right)\mathbf{P}_{11}\mathbf{e} - \left(\frac{\mathbf{e}^T\mathbf{P}_{11}\mathbf{e}}{\mathcal{D}}\right)\mathbf{P}_{12}\mathbf{e},\\ \mathbf{a}_{22} &= \left(\frac{\mathbf{e}^T\mathbf{P}_{12}\mathbf{e}}{\mathcal{D}}\right)\mathbf{P}_{21}\mathbf{e} - \left(\frac{\mathbf{e}^T\mathbf{P}_{11}\mathbf{e}}{\mathcal{D}}\right)\mathbf{P}_{22}\mathbf{e}.\end{aligned} \tag{31}$$

The MMSE associated with the optimal coefficients (31) is $(\varepsilon_{\text{loc}})_{\min} = \frac{\mathbf{e}^T\mathbf{P}_{22}\mathbf{e}+\mathbf{e}^T\mathbf{P}_{11}\mathbf{e}}{\mathcal{D}}$.

In the general p-dimensional case, the Lagrangian function to be minimized has the following form:

$$\begin{aligned}&\Lambda(\mathbf{a}_{ji}, \lambda_{ji}; \quad j, i = 1, \dots, p)\\ &= \sum_{i=1}^{p}\mathbf{a}_{(i)}^T\check{\mathbf{R}}_p\mathbf{a}_{(i)}\\ &\quad + \sum_{j=1}^{p}\left\{\lambda_{jj}(1 - \mathbf{e}^T\mathbf{a}_{jj}) - \sum_{i=1 i\neq j}^{p}\lambda_{ji}\mathbf{e}^T\mathbf{a}_{ji}\right\}.\end{aligned} \tag{32}$$

Differentiating $\Lambda(\mathbf{a}_{ji}, \lambda_{ji}; j, i = 1, \dots, p)$ with respect to $\mathbf{a}_{ji}$ and equating the partial derivatives with zero, p independent sets of equations result, i.e.

$$\check{\mathbf{R}}_p\mathbf{a}^*_{(i)} = \frac{1}{2}\begin{bmatrix}\lambda_{1i}\mathbf{e}\\ \lambda_{2i}\mathbf{e}\\ \vdots\\ \lambda_{pi}\mathbf{e}\end{bmatrix} \quad i = 1, \dots, p \tag{33}$$

which give the optimal coefficients $\mathbf{a}^*_{(i)}$ in terms of the Lagrange multipliers $\lambda_{1i}, \dots, \lambda_{pi}$. Let us assume that $\check{\mathbf{R}}_p^{-1}$ exists and can be decomposed as follows:

$$\check{\mathbf{R}}_p^{-1} = \begin{bmatrix}\mathbf{P}_{11} & \cdots & \mathbf{P}_{1p}\\ \mathbf{P}_{21} & \cdots & \mathbf{P}_{2p}\\ \vdots & \ddots & \vdots\\ \mathbf{P}_{p1} & \cdots & \mathbf{P}_{pp}\end{bmatrix} \tag{34}$$

where $\mathbf{P}_{ij}, \quad i, j = 1, \dots, p$ are $(N \times N)$ square matrices. The Lagrange multipliers $\lambda_{1i}, \dots, \lambda_{pi}$ are obtained by solving the following set of equations:

$$\begin{bmatrix}\mathbf{e}^T\mathbf{P}_{11}\mathbf{e} & \mathbf{e}^T\mathbf{P}_{12}\mathbf{e} & \cdots & \mathbf{e}^T\mathbf{P}_{1p}\mathbf{e}\\ \vdots & & & \vdots\\ \mathbf{e}^T\mathbf{P}_{i1}\mathbf{e} & \mathbf{e}^T\mathbf{P}_{i2}\mathbf{e} & \cdots & \mathbf{e}^T\mathbf{P}_{ip}\mathbf{e}\\ \vdots & & & \vdots\\ \mathbf{e}^T\mathbf{P}_{p1}\mathbf{e} & \mathbf{e}^T\mathbf{P}_{p2}\mathbf{e} & \cdots & \mathbf{e}^T\mathbf{P}_{pp}\mathbf{e}\end{bmatrix}\begin{bmatrix}\lambda_{1i}\\ \vdots\\ \lambda_{ii}\\ \vdots\\ \lambda_{pi}\end{bmatrix} = \begin{bmatrix}0\\ \vdots\\ 2\\ \vdots\\ 0\end{bmatrix} \tag{35}$$

or equivalently $\lambda_{ji} = 2\frac{c_{ij}(\mathbf{G}_p)}{\det(\mathbf{G}_p)}$, where $\mathbf{G}_p = \{G_{ij}\}; i, j = 1, \dots, p$ is the left-hand side $(p \times p)$ square matrix of (35), and $c_{ij}(\mathbf{G}_p)$ stands for the cofactor of the ij element of $\mathbf{G}_p$. In the following, the subscript p will be dropped out for notational simplicity. By substituting into (33), the optimal coefficients of the location-invariant p-channel marginal L filter are obtained, i.e.

$$\mathbf{a}^*_{(i)} = \frac{1}{\det(\mathbf{G})}\check{\mathbf{R}}_p^{-1}\begin{bmatrix}c_{i1}(\mathbf{G})\mathbf{e}\\ c_{i2}(\mathbf{G})\mathbf{e}\\ \vdots\\ c_{ip}(\mathbf{G})\mathbf{e}\end{bmatrix} \quad i = 1, \dots, p \tag{36}$$

and the associated MMSE is given by

$$(\varepsilon_{\text{loc}})_{\min} = \frac{1}{\det^2(\mathbf{G})}\sum_{i=1}^{p}\sum_{j=1}^{p}\sum_{l=1}^{p}c_{ij}(\mathbf{G})c_{il}(\mathbf{G})G_{jl}. \tag{37}$$

It is known that $\sum_{l=1}^{p} G_{jl}c_{il}(\mathbf{G}) = \delta_{ji}\det(\mathbf{G})$ [18], where δ_{ji} denotes Kronecker's delta. Therefore

$$(\varepsilon_{\text{loc}})_{\min} = \frac{1}{\det(\mathbf{G})}\sum_{i=1}^{p}c_{ii}(\mathbf{G}). \tag{38}$$

It is seen that the optimal coefficients (31) or (36) are independent of the two-channel signal to be estimated. Unfortunately, the location-invariant two-channel marginal L filter leads only to a slightly higher noise suppression than its single-channel counterparts, as will be seen later on.

D. Practical Considerations

As can be seen in the preceding analysis, the following difficulties are met in the design of the unconstrained and the unbiased multichannel L filters: i) The marginal L filter coefficients depend explicitly on the signal $\mathbf{s}$ (to be estimated). ii) The marginal and joint probability density functions of the input vector-valued observations must be known in order to calculate $\hat{\mathbf{R}}_p$ and $\hat{\underline{\mu}}_p$. In the following, the estimation of $\mathbf{s}$ is treated both for a multichannel constant signal corrupted by additive white multivariate noise as well as for a nonconstant one. The design of unconstrained and unbiased multichannel marginal L filters based on estimates of the marginal and joint probability density functions of the components of $\mathbf{x}(k)$ is considered as well.

Let us consider first the case of a constant signal $\mathbf{s}$. We restrict the discussion in 1-D signals without any loss of generality. In this case, an initial estimate $\hat{\mathbf{s}}(k)$ for $\mathbf{s}$ could be the arithmetic mean of the past L filter outputs $\mathbf{y}(l)$, $l = k-1, k-2, \ldots$, i.e., $\hat{\mathbf{s}}(k) = \frac{1}{N_e}\sum_{l=k-N_e}^{k-1} \mathbf{y}(l)$, where N_e is chosen to be sufficiently large. The initial estimate $\hat{\mathbf{s}}(k)$ can be used in the place of $\mathbf{s}$ in (13) and (15) or (20) and (21) to determine the multichannel marginal L filter coefficients at each time instant k, provided that $\hat{\mathbf{R}}_p^{-1}$ and $\hat{\underline{\mu}}_p$ have already been computed based on the knowledge of the distributional model of input vector-valued observations. For $k < N_e$, the marginal median can be used to initialize filtering. An even better estimate $\hat{\mathbf{s}}(k)$ can be obtained by employing the arithmetic mean or the marginal median either of the N_e past input vector-valued observations (i.e., noisy observations) or of window size N_e that is centered on the sample $\mathbf{x}(k)$.

A segmentation of a multichannel nonconstant signal (e.g., a multichannel edge) to homogeneous regions where the signal $\mathbf{s}(k)$ is locally constant (i.e., edges do not occur) and to transition regions where an edge occurs in a certain input channel by using an edge detection algorithm (e.g., the one described in [19]) is proposed. We may then use one of the previously described techniques to estimate the constant multichannel signal within each homogeneous region by restricting either the past L filter outputs or the past noisy input vector-valued observations to lie within the homogeneous region.

Next, the design of multichannel marginal L filters based on estimates of the marginal and joint pdf of input vector-valued observations is considered. Both in the case of a multichannel constant signal corrupted by additive white multivariate noise as well as within the homogeneous regions in the case of a noisy nonconstant multichannel signal, the proposed filters will operate on identically distributed observations. Therefore, we can estimate the marginal statistics of input vector data from the empiric pdf, i.e., by uniformly quantizing the range of input observations in each channel to a number, say M, of discrete values and constructing their histogram. Moreover, the input joint statistics can be estimated from the empiric joint pdf, i.e., by exploiting the uniform quantization of any couple of input vector data components to a set of pairs of discrete values and by estimating their cooccurrence matrix. As can be seen in Section V, such an estimation procedure is found to be very successful. Consequently, for homogeneous regions, we may proceed to the calculation of $\hat{\mathbf{R}}_p$ and $\hat{\underline{\mu}}_p$ as described in Section III-B by using the formulae for i.i.d. input variates.

It remains to be examined what actions are taken in the transition regions, where an edge occurs in a certain channel. In the transition regions, input vector-valued observations may be considered to a first approximation as independent nonidentically distributed random variables. Moreover, our problem formulation leads to the design of a filter bank where several different filters should be designed to cope with each instance of the running filter window. However, the formulae for calculating $\hat{\mathbf{R}}_p$ and $\hat{\underline{\mu}}_p$ for independent i.i.d. input variates do not hold anymore in the design of the (unconstrained/unbiased) multichannel signals that fall into the transition region. Even if a least squares criterion were invoked in order to reduce the number of multichannel L filters, as is proposed in the single channel case [5], the framework for calculating $\hat{\mathbf{R}}_p$ and $\hat{\underline{\mu}}_p$ for independent nonidentically distributed input variates is very complicated. In this paper, we shall confine ourselves only to a brief discussion on the the theoretical treatment of multichannel nonconstant signals, and we shall employ the marginal median for filtering the input vector data that belong to transition regions.

III. Computation of the Moments of the Multivariate-Order Statistics

In the previous section, it has been shown that the design of the optimal multichannel marginal L filters depends on the calculation of the composite matrix $\hat{\mathbf{R}}_p$ and the composite vector $\hat{\underline{\mu}}_p$. The diagonal submatrices of $\hat{\mathbf{R}}_p$, $\mathbf{R}_{jj}$, $j = 1, \ldots, p$ consist of moments of the order statistics from a univariate population. The off-diagonal submatrices of $\hat{\mathbf{R}}_p$, $\mathbf{R}_{ji}$, $j = 1, \ldots, p; i \neq j$ consist of moments of the order statistics from a bivariate population. The moments of the order statistics are referred to input signal vector components except in the location-invariant solution. In that case, the moments of the order statistics of input noise vector components are needed. The vector $\hat{\underline{\mu}}_p$ is necessary only in the unconstrained and unbiased solution. Its elements $\underline{\mu}_j$, $j = 1, \ldots, p$ consist of first-order moments of the order statistics from a univariate population.

A discrete calculation of the correlation matrices and mean vectors is described in this section. Such a discrete calculation is needed in order to avoid the extensive numerical integration involved in the definition of the moments of the order statistics [13]. For i.i.d. variates whose distributional model is known, the calculation can be based on the optimal quantization of each input signal (noise) vector component in the mean-squared-error sense. Thus, in the univariate case, a discrete calculation of the moments of the order statistics can be devised by mapping each component of the input vector, which is a continuous random variable, into a discrete random variable by employing the optimum Lloyd–Max quantizer [20], [21]. Similarly, a discrete calculation of the moments of the order statistics in the bivariate case is also needed. To do so, the 2-D vectors of continuous random variables for all possible pairwise combinations between different channels should be mapped into 2-D vectors of discrete random variables by employing vector quantization. In such an approach, numerical integration is required only in the design of Lloyd–Max quantizer, as will be seen later on. For independent nonidentically distributed variates as well as when the distributional model of the input vector-valued observations is unknown, we may relax the requirement for optimal quantization by confining ourselves to uniformly quantizing each input signal vector component. Two-dimensional vector quantization is treated first. Next, we proceed to the calculation of correlation matrices and mean vectors for i.i.d. input observations. However, for the sake of completeness, the calculation of correlation matrices and mean vectors for independent nonidentically distributed variates using inclusion-exclusion identities is briefly discussed.

A. 2-D Vector Quantization

Let us assume the following 2-D vector $\mathbf{x} = (x_1, x_2)^T$ of continuous random variables x_1 and x_2 without any loss of generality. Let $T_1 = \{t_{1k}, k = 1, \ldots, M_1 + 1\}$ and $T_2 = \{t_{2k}, k = 1, \ldots, M_2 + 1\}$ be the sets of increasing decision levels along each dimension where t_{i1} and $t_{i,M_i+1}, i = 1, 2$ denote the minimum and maximum values of x_i, respectively.

Definition 5: A 2-D (M_1, M_2)-level vector quantizer (VQ) maps a vector $\mathbf{x} = (x_1, x_2)^T$ of continuous random variables x_1 and x_2 into a vector $\mathbf{x}_d = (x_1^*, x_2^*)^T$ of discrete random variables x_1^*, x_2^* such that the pair (x_1^*, x_2^*) takes values from the Cartesian product $U = U_1 \times U_2 = \{v_{1k}, k = 1, \ldots, M_1\} \times \{v_{2k}, k = 1, \ldots, M_2\}$ by using the rule

$$\begin{aligned}\mathbf{x}_d &= Q(\mathbf{x}) \\ &= (v_{1m}, v_{2n})^T \\ &\quad \text{if } (t_{1,m} \le x_1 < t_{1,m+1} \text{ and } t_{2,n} \le x_2 < t_{2,n+1})\end{aligned} \tag{39}$$

where $1 \le m \le M_1$ and $1 \le n \le M_2$.

The optimum quantizer minimizes the MSE for given numbers of quantization levels M_1, M_2. Let $f_{x_1,x_2}(x_1, x_2)$ be the joint pdf of the continuous random variables x_1, x_2. It is desired to find the decision levels $t_{1,m}, t_{2,n}$ and the discrete values $v_{1,m}, v_{2,n}$ for an (M_1, M_2)-level VQ such that the MSE

$$\begin{aligned}\varepsilon &= \mathrm{E}[\|\mathbf{x} - \mathbf{x}_d\|^2] \\ &= \sum_{m=1}^{M_1} \sum_{n=1}^{M_2} \int_{t_{1,m}}^{t_{1,m+1}} \int_{t_{2,n}}^{t_{2,n+1}} [(x_1 - v_{1,m})^2 + (x_2 - v_{2,n})^2] \\ &\quad \times f_{x_1,x_2}(x_1, x_2) dx_1 dx_2\end{aligned} \tag{40}$$

is minimized.

Lemma 1: If $t_{1,M_1+1} = -t_{1,1} = +\infty$ and $t_{2,M_2+1} = -t_{2,1} = +\infty$, then the optimal 2-D (M_1, M_2)-level VQ is equivalent to two marginal Lloyd–Max quantizers operating on each dimension independently and designed by using the marginal pdfs, that is, the decision levels $t_{1,m}$, $m = 2, \ldots, M_1 + 1$ and $t_{2,n}$, $n = 2, \ldots, M_2 + 1$ are given, respectively, by

$$t_{1,m} = \frac{v_{1,m} + v_{1,m-1}}{2} \quad t_{2,n} = \frac{v_{2,n} + v_{2,n-1}}{2} \tag{41}$$

and the optimal discrete values $v_{1,m}$, $m = 1, \ldots, M_1$ and $v_{2,n}$, $n = 1, \ldots, M_2$ are obtained, respectively, as follows:

$$\begin{aligned}v_{1,m} &= \frac{\int_{t_{1,m}}^{t_{1,m-1}} x_1 f_{x_1}(x_1) dx_1}{\int_{t_{1,m}}^{t_{1,m+1}} f_{x_1}(x_1) dx_1}, \\ v_{2,n} &= \frac{\int_{t_{2,n}}^{t_{2,n+1}} x_2 f_{x_2}(x_2) dx_2}{\int_{t_{2,n}}^{t_{2,n-1}} f_{x_2}(x_2) dx_2}.\end{aligned} \tag{42}$$

Therefore, the 2-D vector quantization degenerates into two independent quantizations along each dimension and can be done at no additional cost since 1-D quantization is necessary to the calculation of the moments of the order statistics in the univariate case. In other words, there exists a unified framework for the discrete calculation of the moments of the order statistics in both the univariate and bivariate case. m-D vector quantization is not treated because moments of the order statistics from a bivariate population are required only in the design of the multichannel marginal L filters described in Section II.

B. Calculation of Correlation Matrices and Mean Vectors for i.i.d. Input Observations

This section is devoted to the calculation of the moments of the order statistics of the signal vector components for i.i.d. input observations. The same procedure can also be applied to the calculation of the moments of the order statistics of the noise vector components, if the signal components are replaced by the noise components.

Let us assume that x_j, $j = 1, \ldots, p$ has been quantized to M discrete values, i.e., the discrete random variable $x_j^* \in U_j = \{v_{j,1}, v_{j,2}, \ldots, v_{j,M}\}$. The elements of any submatrix of $\hat{\mathbf{R}}_p$, say $\mathbf{R}_{ji}$, $j, i = 1, \ldots, p$ are given by

$$\mathbf{R}_{ji}^{rs} = \mathrm{E}[x_{j(r)} x_{i(s)}] = \sum_{m=1}^{M} \sum_{n=1}^{M} v_{j,m} v_{i,n} f_{(r,s;j,i)}(v_{j,m}, v_{i,n}) \qquad r, s = 1, \ldots, N; j, i = 1, \ldots, p \tag{43}$$

$$\begin{aligned}f_{(r,s;j,i)}(v_{j,m}, v_{i,n}) &\stackrel{\text{def}}{=} \Pr\{x_{j(r)} = v_{j,m}, x_{i(s)} = v_{i,n}\} \\ &= \begin{cases} F_{(r,s;j,i)}(v_{j,m}, v_{i,n}) + F_{(r,s;j,i)}(v_{j,m-1}, v_{i,n-1}) & \\ -F_{(r,s;j,i)}(v_{j,m}, v_{i,n-1}) - F_{(r,s;j,i)}(v_{j,m-1}, v_{i,n}) & m = 2, \ldots, M \quad n = 2, \ldots, M \\ F_{(r,s;j,i)}(v_{j,1}, v_{i,n}) - F_{(r,s;j,i)}(v_{j,1}, v_{i,n-1}) & m = 1 \quad n = 2, \ldots, M \\ F_{(r,s;j,i)}(v_{j,m}, v_{i,1}) - F_{(r,s;j,i)}(v_{j,m-1}, v_{i,1}) & m = 2, \ldots, M \quad n = 1 \\ F_{(r,s;j,i)}(v_{j,1}, v_{i,1}) & m = 1 \quad n = 1. \end{cases}\end{aligned} \tag{44}$$

where we have (44), which appears at the bottom of the previous page. For $i = j$, $F_{(r,s;j,j)}(v_{j,m}, v_{j,n})$ denotes the value at $(v_{j,m}, v_{j,n})$ of the cumulative distribution function (cdf) of the order statistics from a univariate population, and the well-known formulae [13], [1] can be applied:

$$\begin{aligned}F_{(r,s;j,j)}&(v_{j,m}, v_{j,n})\\ &= \sum_{q_2=s}^{N}\sum_{q_1=r}^{q_2} \frac{N!}{q_1!(q_2-q_1)!(N-q_2)!}\\ &\times F_{x_j}^{q_1}(v_{j,m})(F_{x_j}(v_{j,n}) - F_{x_j}(v_{j,m}))^{q_2-q_1}\\ &\times (1 - F_{x_j}(v_{j,n}))^{N-q_2} \quad \text{if} \quad v_{j,m} < v_{j,n} \qquad (45)\end{aligned}$$

$$\begin{aligned}F_{(s;j)}(v_{j,n}) &= \sum_{q=s}^{N} \frac{N!}{q!(N-q)!} F_{x_j}^{q}(v_{j,n})(1 - F_{x_j}(v_{j,n}))^{N-q}\\ &\text{if } v_{j,m} > v_{j,n}. \qquad (46)\end{aligned}$$

The values of $F_{x_j}(v_{j,m})$ at $v_{j,m}$ are calculated in terms of the probabilities of discrete events $\{x_j^* = v_{j,q}\}$, i.e.

$$F_{x_j}(v_{j,m}) = \sum_{q=1}^{m} \Pr\{x_j^* = v_{j,q}\} \quad m = 1, \ldots, M. \qquad (47)$$

The probabilities involved in (47) depend on the optimal decision levels along the jth dimension of the signal determined by the Lloyd–Max quantizer design

$$\Pr\{x_j^* = v_{j,q}\} = \int_{t_{j,q}}^{t_{j,q+1}} f_{x_j}(x_j)dx_j \qquad (48)$$

where $f_{x_j}(x_j)$ denotes the marginal pdf of the jth random variable x_j.

Next, we shall proceed to the calculation of the cdf values for $i \neq j$, i.e., the cdf values of the 2-D order statistic $(x_{j(s)}, x_{i(r)})$ at $(v_{j,m}, v_{i,n})$. To do so, the results reported in pp. 25 and 26: Exercise 2.2.2 of [13] and [7] and [8] will be exploited. More specifically, it has been proved

$$\begin{aligned}F_{(r,s;j,i)}&(v_{j,m}, v_{i,n}) \overset{\text{def}}{=} \Pr\{x_{j(r)} \leq v_{j,m}, x_{i(s)} \leq v_{i,n}\}\\ &= \sum_{q_1=r}^{N}\sum_{q_2=s}^{N}\sum_{\rho=\max(0,q_1+q_2-N)}^{\min(q_1,q_2)}\\ &\times \frac{N!}{\rho!(q_1-\rho)!(q_2-\rho)!(N-q_1-q_2+\rho)!}\\ &\times \mathcal{F}_0^{\rho}(v_{j,m}, v_{i,n})\mathcal{F}_1^{q_2-\rho}(v_{j,m}, v_{i,n})\mathcal{F}_2^{q_1-\rho}(v_{j,m}, v_{i,n})\\ &\times \mathcal{F}_3^{N-q_1-q_2+\rho}(v_{j,m}, v_{i,n}) \qquad (49)\end{aligned}$$

where $\mathcal{F}_k(v_{j,m}, v_{j,n})$, $k = 0, \ldots, 3$ are the probability masses in the four regions of the (x_j, x_i) plane defined by the following equations:

$$\begin{aligned}\mathcal{F}_0(v_{j,m}, v_{i,n}) &= \Pr\{x_j \leq v_{j,m}, x_i \leq v_{i,n}\}\\ &= F_{x_j,x_i}(v_{j,m}, v_{i,n})\\ \mathcal{F}_1(v_{j,m}, v_{i,n}) &= \Pr\{x_j > v_{j,m}, x_i \leq v_{i,n}\}\\ \mathcal{F}_2(v_{j,m}, v_{i,n}) &= \Pr\{x_j \leq v_{j,m}, x_i > v_{i,n}\}\\ \mathcal{F}_3(v_{j,m}, v_{i,n}) &= \Pr\{x_j > v_{j,m}, x_i > v_{i,n}\} \qquad (50)\end{aligned}$$

and $F_{x_j,x_i}(v_{j,m}, v_{i,n})$ is the value of the joint cdf of random variables x_j, x_i at $(v_{j,m}, v_{i,n})$. The above-defined probability masses can be easily calculated in terms of the probabilities of the discrete events $\{x_j^* = v_{j,m}, x_i^* = v_{i,n}\}$. The computation of these probabilities depends on the decision levels along the jth and ith dimensions provided by the Lloyd–Max quantizers and can be done without any difficulty.

The calculation of the elements of each mean vector $\underline{\mu}_j$, $j = 1, \ldots, p$ is performed by using

$$\mathrm{E}[x_{j(r)}] = \sum_{m=1}^{M} v_{j,m} f_{(r;j)}(v_{j,m}) \; r = 1, \ldots, N \qquad (51)$$

where

$$\begin{aligned}f_{(r;j)}&(u_{j,m})\\ &= \begin{cases} F_{(r;j)}(v_{j,m}) - F_{(r;j)}(v_{j,m-1}) & m = 2, \ldots, M \\ F_{(r;j)}(v_{j,1}) & m = 1\end{cases} \qquad (52)\end{aligned}$$

and $F_{(r;j)}(v_{j,m})$ are calculated by using (46).

C. Calculation of Correlation Matrices and Mean Vectors for Independent Nonidentically Distributed Observations

For independent nonidentically distributed observations, (45), (46), and (49) no longer hold. Two approaches can be found in the literature for the calculation of the first- and second-order moments of the order statistics from a univariate population. The first approach exploits the notion of the *permanent* of a matrix and is proposed by Vaughan and Venables [22]. Descriptions of that approach can also be found in [1] and [5]. The second one is proposed by Maurer and Margolin [23]. It is based on multivariate inclusion-exclusion identities. It has been used successfully in the development of recursive discrete algorithms for computing the marginal and joint cumulative function (cdf) of the order statistics from a univariate population in polynomial time [3]. A similar approach based on inclusion-exclusion identities has also been proposed for the computation of the joint cumulative function of the order statistics from a bivariate population [24], [25].

IV. Multivariate Distributions

Single-channel L filters have been proved efficient for nonGaussian, long-tailed, or short-tailed noise filtering. The previous attempts to use nonlinear filters based on order statistics for vector-valued signal processing have been derived either from a natural generalization of univariate exponential distributions [6] (that are not Laplacian as explained later on) or have been tested on a contaminated multinormal distribution that has been used to model long-tailed multivariate distributions [7]–[9]. Therefore, the need emerges for the design of long-tailed multivariate distributions, especially Laplacian ones, in order to test the performance of nonlinear filters based on order statistics for vector-valued signal restoration. The multichannel marginal L filters examined in the present paper are designed to be used for nonGaussian bivariate noise filtering.

Let us define first the notion of the uniform, Gaussian, Laplacian, etc. multivariate distributions.

Definition 6 [26], [28]: A multivariate distribution is said to be uniform, Gaussian, Laplacian, etc. when the univariate marginal distributions are all uniform, Gaussian, Laplacian, etc.

It is known that a distribution of the form $f(\mathbf{x}) = \gamma \exp[-\alpha\|\mathbf{x} - \underline{\theta}\|_2]$, where $\| \ \|_2$ denotes the L_2 norm, has led to the definition of the vector median using the L_2 norm (VM_{L_2}). Although, the authors claim that such a distribution is a bivariate biexponential distribution, in fact, it is a generalization of the so-called Subbotin univariate distribution, and its univariate marginal distributions are not Laplacian [26]. Consequently, this distribution is not Laplacian according to Definition 6. In the following, the design of a bivariate Laplacian distribution is examined.

Long-tailed elliptically symmetric distributions can be found in the literature [26]. They have density functions depending only on quadratic functions of the variables. We shall follow a different approach in the design of a bivariate Laplacian joint pdf [26], [28]. The derivation of a bivariate Laplacian distribution is based on Lemma 2.

Lemma 2: A joint distribution $F_{x_1,x_2}(x_1,x_2)$ given by:

$$F_{x_1,x_2}(x_1,x_2) = F_{x_1}(x_1)F_{x_2}(x_2)[1+\alpha(1-F_{x_1}(x_1)) \times (1-F_{x_2}(x_2))] \tag{53}$$

where $\alpha \in [-1,+1]$ has as marginal cdf's $F_{x_i}(x_i) i = 1,2$.

The proof of Lemma 2 can be found in [26] and [27]. The family of joint distributions (53) is the so-called Morgenstern's family. We are interested in the following case:

$$F_{x_i}(x_i) = \begin{cases} \frac{1}{2}\exp[\sqrt{2}\frac{x_i}{\sigma_i}] & \text{if } x_i < 0 \\ 1 - \frac{1}{2}\exp[-\sqrt{2}\frac{x_i}{\sigma_i}] & \text{if } x_i \geq 0 \end{cases} \quad i = 1,2 \tag{54}$$

which results in the bivariate Laplacian distribution, which belongs to the Morgenstern family. The following lemma determines the joint pdf of the bivariate Laplacian distribution, which belongs to the Morgenstern family and relates the correlation coefficient r to the parameter α.

Lemma 3:

a) The joint pdf of the bivariate Laplacian distribution that belongs to the Morgenstern family is given by

$$\begin{aligned} f_{x_1,x_2}&(x_1,x_2) \\ &= \frac{1}{2\sigma_1\sigma_2}\exp\left[-\sqrt{2}\left(\frac{|x_1|}{\sigma_1}+\frac{|x_2|}{\sigma_2}\right)\right] \\ &\times\left\{1+\alpha \ \mathrm{sgn}(x_1,x_2)\left(1-\exp\left[-\sqrt{2}\frac{|x_1|}{\sigma_1}\right]\right)\right. \\ &\left.\times\left(1-\exp\left[-\sqrt{2}\frac{|x_2|}{\sigma_2}\right]\right)\right\} \end{aligned} \tag{55}$$

where

$$\mathrm{sgn}(x_1,x_2) = \begin{cases} 1 & x_1x_2 \geq 0 \\ -1 & x_1x_2 < 0. \end{cases} \tag{56}$$

b) Let $f_{x_1,x_2}(x_1,x_2;\theta_1,\theta_2) = f_{x_1,x_2}(x_1-\theta_1, x_2-\theta_2)$, where $\underline{\theta} = (\theta_1,\theta_2)^T$ is the location parameter of the distribution. The correlation coefficient r for the bivariate Laplacian distribution under study is given by $r = \frac{9}{32}\alpha$, i.e., the correlation coefficient cannot exceed $9/32$.

The proof of Lemma 3 is straightforward. Therefore, it is omitted. Next, we shall give the definition of the generalized vector median of N vector-valued observations:

Definition 7 [6]: The generalized vector median (GVM) of $\mathbf{x}_1,\ldots,\mathbf{x}_N$ is the vector $\mathbf{x}_{\mathrm{gvm}} \in \{\mathbf{x}_1,\ldots,\mathbf{x}_N\}$ such that $\forall j,\ j=1,\ldots,N \ \sum_{i=1}^N \mathrm{dist}(\mathbf{x}_{\mathrm{gvm}},\mathbf{x}_i) \leq \sum_{i=1}^N \mathrm{dist}(\mathbf{x}_j,\mathbf{x}_i)$, where $\mathrm{dist}(\mathbf{x}_j,\mathbf{x}_i)$ is a distance function between the vectors $\mathbf{x}_j$ and $\mathbf{x}_i$.

Using Definition 7, the maximum likelihood estimator (MLE) of location for the Laplacian–Morgenstern distribution is determined by the following lemma:

Lemma 4: For $\sigma_1 = \sigma_2 = \sqrt{2}$, the MLE of the location vector $\underline{\theta}$ based on a random sample of N observations $\{\mathbf{x}_1,\ldots,\mathbf{x}_N\}$, where $\mathbf{x}_i = (x_{i1},x_{i2})^T i = 1,\ldots,N$, is the generalized vector median (GVM) that uses the following distance function between the vectors $\mathbf{x}_j$ and $\mathbf{x}_i$:

$$\begin{aligned} \mathrm{dist}&(\mathbf{x}_j,\mathbf{x}_i) \\ &= \|\mathbf{x}_j-\mathbf{x}_i\|_1 - \ln[1+\alpha \ \mathrm{sgn}(x_{j1}-x_{i1},x_{j2}-x_{i2}) \\ &\quad \times (1-\exp[-|x_{j1}-x_{i1}|])(1-\exp[-|x_{j2}-x_{i2}|])]. \end{aligned} \tag{57}$$

The proof of Lemma 4 is simple and is omitted. An interesting conclusion can be deduced from (57). If the components of the vector-valued signal are uncorrelated, i.e., $\alpha = 0$, the distance between any two vectors is reduced to the L_1 norm, and the vector median using the L_1 norm results. This conclusion has also been derived in [6], starting from a different origin.

We conclude the discussion on the bivariate Laplacian distribution with the design of a noise generator for the abovementioned distribution. The design is based on Lemma 5.

Lemma 5: Let $\mathbf{z} = (z_1,z_2)^T$ be bivariate uniform with joint density $g_{z_1,z_2}(z_1,z_2)$ given by

$$g_{z_1,z_2}(z_1,z_2) = 1+\alpha(2z_1-1)(2z_2-1). \tag{58}$$

Let $f_1(x_1), f_2(x_2)$ be univariate Laplacian pdf's, i.e.

$$f_i(x_i) = \frac{1}{\sigma_i\sqrt{2}}\exp\left[-\sqrt{2}\frac{|x_i|}{\sigma_i}\right], \quad i=1,2 \tag{59}$$

with corresponding cdf's (54) $F_1(x_1)$ and $F_2(x_2)$, respectively. Then, the joint pdf of $\mathbf{x} = (x_1,x_2)^T = (F_1^{-1}(z_1),F_2^{-1}(z_2))^T$ is bivariate Laplacian (55).

Let z_1,u,v be independent uniformly distributed in the interval $[0,1]$ random variables. If z_2 is produced as follows:

$$z_2 = \begin{cases} \min\left(u, \frac{-v}{\alpha(2z_1-1)}\right) & z_1 < 1/2 \\ \max\left(u, 1-\frac{v}{\alpha(2z_1-1)}\right) & z_1 \geq 1/2 \end{cases} \tag{60}$$

the random vector $\mathbf{z} = (z_1,z_2)^T$ is distributed according to (58). Consequently, a random vector $\mathbf{x} = (x_1,x_2)^T$, which has bivariate Laplacian distribution $f_{x_1,x_2}(x_1,x_2;\theta_1,\theta_2)$ belonging to the Morgenstern family can be obtained by

$$x_i = \begin{cases} \theta_i - \frac{\sigma_i}{\sqrt{2}}\ln[2(1-z_i)] & z_i \geq 1/2 \\ \theta_i + \frac{\sigma_i}{\sqrt{2}}\ln[2z_i] & z_i < 1/2 \end{cases} \quad i=1,2. \tag{61}$$

V. Simulation Examples

In this section, we present two sets of experiments in order to test the performance of the multichannel marginal L filters we have designed in Section II. In the first set of experiments, two-channel 1-D input signal sequences generated by corrupting either a constant signal or an edge by additive white bivariate noise have been used. In the second set of experiments, we deal with two-channel 2-D artificially generated vector fields. These fields have been produced by corrupting either a constant velocity field or a velocity field exhibiting edges by additive white Laplacian–Morgenstern bivariate noise.

A. Two-Channel 1-D Signals

First, the case of a two-channel 1-D constant signal $\mathbf{s}(k) = \mathbf{s}$ corrupted by additive white bivariate noise obeying various noise models is treated. The following noise models have been considered: joint Gaussian, uniform (58), contaminated Gaussian, and Laplacian–Morgenstern. The performance of the unconstrained, unbiased, and location-invariant multichannel marginal L filters in multichannel noise filtering has been compared with the performance of other multichannel nonlinear filters as well as their single -channel counterparts. The following nonlinear filters have been considered: the vector median [6], the marginal median [7], [8], the marginal α-trimmed mean [7], [8], the multichannel modified trimmed mean (MTM) [8], the multichannel double window modified trimmed mean (DW-MTM) [8] and the ranked-order estimator $\mathcal{R}_E$ [9]. The multichannel DW-MTM filter uses two window sizes, as in the single channel case [30]. The small window size $N_s = 2\nu_s + 1$ is chosen to be seven for $N = 9$. Its output is determined as follows. First, the marginal median $\mathbf{x}_{\nu_s+1}$ is computed inside the small window. Then, the data vectors lying in the large window of size $N = 2\nu + 1$ having distances $d(l) = (\mathbf{x}(l) - \mathbf{x}_{\nu_s+1})^T(\mathbf{x}(l) - \mathbf{x}_{\nu_s+1})$, $l = k - \nu, \ldots, k + \nu$ less than the average distance $\overline{d}(k)$ are averaged to give the filter output. The multichannel MTM filter is a generalization of the single-channel one proposed in [30]. Its output is evaluated as for the DW-MTM filter, except that only one window of size N is used. The trimming parameter for α-trimmed mean filters has been 0.2 in all experiments. For $\mathcal{R}_E$ filter [9], only the best result is tabulated. It has been found that the highest NR has been obtained for $j = 1$ in all experiments with constant signals corrupted by additive white bivariate noise. Since the concept of radial median is better suited for smoothing directional data (e.g., angular data) [29], multichannel L filters based on radial medians [10] will not be considered in our comparative study. We have also included the arithmetic mean in the comparative study because it is a straightforward choice for noise filtering in many practical applications. The performance of the single-channel L filter counterparts, i.e., the unbiased and location-invariant single-channel L filter [1] used to filter the noise in each channel independently, has also been taken into consideration. The quantitative criterion we used was the noise reduction index (NR) defined as the ratio of the output noise power to the input noise power, i.e.

$$\mathrm{NR} = 10 \log \frac{\sum_k (\mathbf{y}(k) - \mathbf{s})^T(\mathbf{y}(k) - \mathbf{s})}{\sum_k (\mathbf{x}(k) - \mathbf{s})^T(\mathbf{x}(k) - \mathbf{s})}. \tag{62}$$

First, we shall assume that $\mathbf{s}(k) = \mathbf{s}$ is known and, furthermore, that the distributional model of the input vector-valued observations is also known. Next, we shall examine quantitatively the effectiveness of the estimation procedures for $\mathbf{s}$, $\hat{\mathbf{R}}_2$, and $\underline{\hat{\mu}}_2$ developed in Section II-D. Finally, the treatment of a two-channel 1-D edge is discussed. In the tables following, the ranking of each multichannel nonlinear filter for each experiment can be found inside brackets.

The performance of the multichannel DW-MTM, MTM, the $\mathcal{R}_E$ estimator, the marginal median, the α-trimmed mean, and the location-invariant single-channel L filters (i.e., arithmetic mean filters) used to filter each channel independently has been compared with the one of the location-invariant two-channel L filter for the joint Gaussian noise. A vector-valued constant signal $\mathbf{s} = (1.0, 1.0)^T$ corrupted by additive white bivariate noise $\mathbf{n}(k)$ whose components are distributed according to the joint Gaussian distribution $\mathcal{N}(m_1 = 0.0, m_2 = 0.0; \sigma_1 = 1.0, \sigma_2 = 3.0;\ r = -0.5)$ has been used as a test signal. The parameters $m_i, \sigma_i, i = 1, 2$ denote the expected value and the standard deviation of each noise vector component, respectively, and r is the correlation coefficient. The NR index is shown in Table I for filter length $N = 5$. It is straightforward to prove that the MLE of the constant signal $\mathbf{s}$ for jointly Gaussian observations is the arithmetic mean filter. As expected, the design of location-invariant two-channel marginal L filter (31) has led to a filter that resembles the MLE in the sense that the elements of the coefficient vectors $\mathbf{a}_{11}$ and $\mathbf{a}_{22}$ are close to $1/N$, and those of $\mathbf{a}_{12}$ and $\mathbf{a}_{21}$ are close to 0. Since the correlation between the components of the input vector-valued observations is not exploited at all, we have not proceeded to the design of the unconstrained or unbiased two-channel marginal L filters.

The performance of the multichannel filters under study has been evaluated for bivariate uniform noise (58). A vector-valued constant signal $\mathbf{s} = (1.0, 2.0)^T$ corrupted by additive white bivariate noise $\mathbf{n}(k)$ whose components are correlated and uniformly distributed in the interval $[-0.5, 0.5]$ along each channel has been used as a test signal. In other words, the joint pdf of the noise vector components $\mathbf{n}(k) = (n_1(k), n_2(k))^T$ is given by

$$f_{n_1,n_2}(n_1, n_2) = 1 + 4\alpha n_1 n_2. \tag{63}$$

In our experiment, $\alpha = 1.0$. The NR index is shown in the first column of Table II for filter length $N = 9$. The case of a contaminated Gaussian noise has been considered as well. A vector-valued constant signal $\mathbf{s} = (1.0, 2.0)^T$ corrupted by additive white bivariate noise $\mathbf{n}(k)$ whose components are distributed according to the contaminated Gaussian distribution given by

$$(1 - \epsilon)\mathcal{N}(0.0, 0.0; 1.0, 3.0; 0.5) + \epsilon\mathcal{N}(0.0, 0.0; 3.0, 9.0; 0.7) \tag{64}$$

for $\epsilon = 0.1$ has been used as a test signal. The NR index is listed in the second column of Table II for filter

TABLE I
NOISE REDUCTION (IN DECIBELS) FOR THE JOINT GAUSSIAN DISTRIBUTION NOISE MODEL (FILTER LENGTH $N = 5$)

Filter	NR
multichannel DW-MTM	-3.337
multichannel MTM	-4.528
$\mathcal{R}_E$	-5.42
marginal median	-5.583
α-trimmed mean	-6.585
location-invariant single-channel L-filter (arithmetic mean)	-7.161
location-invariant 2-channel marginal L-filter	-7.164 √

TABLE II
NOISE REDUCTION (IN DECIBELS) FOR THE BIVARIATE UNIFORM CONTAMINATED GAUSSIAN AND LAPLACIAN-MORGENSTERN DISTRIBUTION NOISE MODELS (FILTER LENGTH $N = 9$)

Filter	NR					
	Uniform		Contaminated Gaussian		Laplacian-Morgenstern	
multichannel DW-MTM	-3.382	[11]	-8.269	[12]	-9.010	[11]
$\mathcal{R}_E$	-5.053	[9]	-8.541	[11]	-8.361	[12]
arithmetic mean	-9.166	[6]	-9.721	[7]	-9.542	[9]
multichannel MTM	-4.490	[10]	-8.821	[9]	-10.39	[8]
vector median L_1	-		-8.767	[10]	-9.01	[11]
generalized vector median	-		-		-9.08	[10]
marginal median	-5.533	[8]	-9.66	[8]	-10.637	[6]
α-trimmed mean	-8.038	[7]	-10.834	[6]	-10.568	[7]
location-invariant single-channel L-filter	-11.81	[5]	-10.882	[5]	-11.087	[5]
location-invariant two-channel marginal L-filter	-11.9	[4]	-10.997	[4]	-11.055	[4]
unbiased single-channel L-filter	-12.212	[3]	-16.365	[3]	-13.548	[3]
unbiased two-channel marginal L-filter	-14.154	[2]	-18.346	[2]	-15.666	[2]
unconstrained two-channel marginal L-filter	-14.154	[1]	-18.564	[1]	-15.728	[1]

length $N = 9$. Next, the case of a vector-valued constant signal $\mathbf{s} = (1.0, 2.0)^T$ corrupted by additive white bivariate noise $\mathbf{n}(k)$ whose components are distributed according to the Laplacian–Morgenstern distribution with zero-mean vector $\sigma_1 = \sigma_2 = \sqrt{2}$ and $\alpha = 1.0$ has been studied. The NR index is tabulated in the third column of Table II for filter length $N = 9$. From the entries of Table II it is clearly seen that the unconstrained two-channel marginal L filter outperforms any other filter. The unbiased two-channel marginal L filter is the second best. The unbiased two-channel L filter attains again almost 2 dB higher noise suppression than its single-channel counterpart. This superior behavior is attributed to the fact that the unbiased two-channel marginal L filter utilizes the correlation between the components of the input vector-valued signal. The performance of the location-invariant two-channel L filter is relatively poor. Furthermore, the proposed unconstrained and unbiased two-channel L filters have better performance than all the other multichannel estimators included in the comparative study. The price paid for the superior performance is the complicated design procedure.

Subsequently, we examine the performance of an unbiased two-channel marginal L filter of length $N = 9$ that is designed by using the estimation procedures for $\mathbf{s}$ developed in Section II-D. We assume that the distributional model of input vector-valued observations is known, and we focus on the estimation of the unknown signal $\mathbf{s}$. Several experiments have been performed on a two-channel 1-D input sequence produced

TABLE III
NOISE REDUCTION (IN DECIBELS) FOR N_e =49, 81 WHEN s IS ESTIMATED BY FILTERING THE NOISY INPUT OBSERVATIONS FOR THE LAPLACIAN-MORGENSTERN DISTRIBUTION NOISE MODEL (FILTER LENGTH $N = 9$)

Window Type	Estimator	NR			
		$N_e = 49$		$N_e = 81$	
		two-channel	single-channel	two-channel	single-channel
causal	arithmetic mean	-12.9486	-11.5043	-13.9665	-12.2742
	marginal median	-13.7109	-12.1142	-14.5033	-12.701
non-causal	arithmetic mean	-12.5191	-10.9499	-13.6084	-11.864
	marginal median	-13.3212	-11.6415	-14.2062	-12.3783
optimal unbiased filter		-15.666	-13.548	-15.666	-13.548

by corrupting a constant signal $\mathbf{s}$ by additive white bivariate Laplacian–Morgenstern noise. In these experiments, we use the noise reduction achieved at the filter output as a figure of merit. Let us first consider that an initial estimate for $\mathbf{s}(k) = \mathbf{s}$ is obtained by averaging the past L filter outputs. It has been found that a very large N_e is required in order to achieve a NR larger than that achieved by employing the marginal median (e.g., $N_e = 225$ for a NR $= -12.309$ dB). Motivated by the extremely large N_e required in the procedure described above in order to avoid instabilities, we have examined the possibility of obtaining an initial estimate for $\mathbf{s}$ from the noisy input vector-valued observations. Two types of estimation procedures have been employed, namely, the estimation of $\mathbf{s}$ in a causal window as well as the estimation of $\mathbf{s}$ in a window centered on the current input data vector $\mathbf{x}(k)$. In the former case, $\mathbf{s}$ is being estimated from the past noisy input observations. In the later case, future noisy input data vectors are also taken into account. A convenient estimator, e.g., the marginal median or the arithmetic mean of length N_e, has been used successfully to provide an estimate $\hat{\mathbf{s}}(k)$ of $\mathbf{s}$. The noise reduction achieved at the filter output for an unbiased two-channel marginal L filter having length $N = 9$ is tabulated for window sizes $N_e = 49$ and $N_e = 81$ in Table III. The corresponding noise reduction when two unbiased single-channel L filters employing the same estimation procedures for $\mathbf{s}$ operate on the two input channels independently is also included in Table III for comparison purposes. It is seen that a much smaller window size N_e is now required. By examining Table III, it is found that causal windows yield better results than noncausal ones and that the marginal median is more efficient in estimating $\mathbf{s}$ than the arithmetic mean, as expected. Furthermore, it is clearly seen that the utilization of the correlation between input components leads to 1.5 dB higher noise suppression. By comparing the entries of Table III with the results tabulated in the third column of Table 2, it is seen that the unbiased two-channel marginal L filter that employs the above-described estimation procedure outperforms all the other multichannel nonlinear filters included in our comparative study.

Next, we consider the case where the distributional model of input vector-valued observations is unknown. In such a case, the moments of the order statistics that form the matrix $\hat{\mathbf{R}}_2$ and the mean vector $\hat{\underline{\mu}}_2$ are calculated by using estimates of the marginal and joint probability density function of input vector data. The design of an unbiased two-channel marginal L filter of length $N = 5$ is treated for the Laplacian–Morgenstern noise model. The deterioration in the filter output due to the

TABLE IV
NOISE REDUCTION (IN DECIBELS) FOR AN UNBIASED TWO-CHANNEL L FILTER OF LENGTH $L = 5$, WHEN $\hat{\mathbf{r}}_2$ AND $\hat{\underline{\mu}}_2$ AND/OR s ARE ESTIMATED

Method	NR		N_e
	two-channel	single-channel	
$\hat{\mathbf{R}}_2$ and $\hat{\underline{\mu}}_2$ are estimated, $\mathbf{s}$ is known	-12.5610	-10.4117	-
$\hat{\mathbf{R}}_2$, $\hat{\underline{\mu}}_2$ and s are estimated	-10.5719	-8.9349	25
	-11.4929	-9.0498	49
optimal unbiased filter	-12.617	-10.494	-

estimation of marginal and joint input pdf is measured first. As can be seen in Table IV, the deterioration in noise suppression is almost negligible (0.056 dB). When $\mathbf{s}$ is estimated from the marginal median of the N_e past input data vectors and $\hat{\mathbf{R}}_2$ and $\hat{\underline{\mu}}_2$ are calculated based on estimates of the marginal and joint pdfs of input data vectors, the total deterioration varies between 1.12 dB for $N_e = 49$ to 2 dB for $N_e = 25$. The same procedure has also been applied with similar success to the design of two unbiased single-channel L filters that have been used to filter the two input channels independently. Again, it is verified experimentally that the unbiased two-channel marginal L filter is superior to its single-channel counterparts, yielding an almost 2 dB higher NR index. We conclude that by combining estimates for $\mathbf{s}$ with $\hat{\mathbf{R}}_2$ and $\hat{\underline{\mu}}_2$ calculated based on estimates of the marginal and joint pdf's of $\mathbf{x}(k)$, the loss in noise-reduction capability of the designed unbiased two-channel marginal L filters is small enough. Moreover, the superiority in noise suppression of the designed unbiased two-channel marginal L filters to the other multichannel estimators as well as to their single channel counterparts still holds.

Finally, the treatment of a two-channel 1-D edge corrupted by additive white Laplacian–Morgenstern bivariate noise is discussed. The case of a bivariate noise with zero-mean vector $\sigma_1 = \sigma_2 = \sqrt{2}$ and $\alpha = 1.0$ is considered. In the first channel, a transition from level 1 to level 10 occurs at time instance k_1. In the second channel, a transition from level 2 to level 15 occurs at time instant $k_2 \neq k_1$. The following edge-detection algorithm has been used [19]. Let W_L, W_R be two neighboring windows, as shown in Fig. 1(a). Let $\hat{x}_{1L}, \hat{x}_{1R}$ be the medians of the samples in the corresponding windows in channel 1. Similarly, let $\hat{x}_{2L}, \hat{x}_{2R}$ denote the medians of the samples in the corresponding windows in channel 2. If

$$|\hat{x}_{1L} - \hat{x}_{1R}| > \tau_1 \text{ or } |\hat{x}_{2L} - \hat{x}_{2R}| > \tau_2, \tag{65}$$

where it is declared that $\mathbf{x}(k)$ belongs to a transition region. In (65), $\hat{x}_{iL}, \hat{x}_{iR}$, $i = 1, 2$ have been evaluated by employing medians of window size 11, and the thresholds τ_i are set equal to 3 (i.e., $\tau_i \simeq 2\sigma_i$). For each homogeneous region, we design the unbiased two-channel marginal L filter of length $N = 5$ that is matched to its statistics by using estimates for both $\hat{\mathbf{R}}_2$, $\hat{\underline{\mu}}_2$ and $\mathbf{s}$ as has been previously described. In the transition regions, we use the marginal median of the same length to filter the data vectors that lie into them. The results for noise reduction are summarized in Table V. The unbiased single-channel L filters that have been included in Table V have also been designed based on estimates for $\hat{\mathbf{R}}_2$, $\hat{\underline{\mu}}_2$, and $\mathbf{s}$. Once more, it is seen that the unbiased two-channel L filter is the best,. yielding an almost 2-dB higher noise suppression.

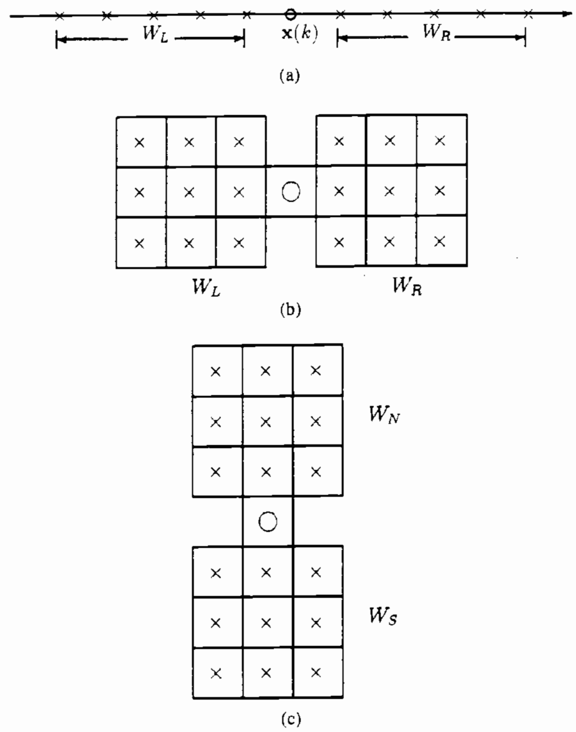


Fig. 1. (a) Windows used in edge detection for 1-D multichannel signals; (b) windows used to detect vertical edges for 2-D multichannel signals; (c) Windows used to detect horizontal edges for 2-D multichannel signals.

TABLE V
NOISE REDUCTION (IN DECIBELS) FOR A TWO-CHANNEL 1-D EDGE CORRUPTED BY ADDITIVE WHITE LAPLACIAN-MORGENSTERN BIVARIATE NOISE (FILTER LENGTH $N = 5$)

Filter	NR
multichannel DW-MTM ($N_s = 3$)	-4.8945
$\mathcal{R}_E$ ($j = 1$)	-5.7939
vector median L_2	-5.7978
vector median L_1	-6.2295
arithmetic mean	-6.9302
multichannel MTM	-7.1135
marginal median	-7.6085
α-trimmed mean	-7.8996
unbiased single-channel L-filter	-8.2951
unbiased two-channel L-filter	-10.2365 √

B. Application to Artificially Generated Velocity Fields

The second set of experiments has been performed on two-channel 2-D artificially generated velocity fields. First, the case of artificially generated vector data that are produced by corrupting a constant velocity vector by additive white Laplacian–Morgenstern bivariate noise is considered. Fig. 2(a) shows a 20×20 block of an artificially generated vector field extracted from the lower-right corner of the original noisy velocity field of dimensions 64×64. The velocity data have been produced by corrupting the constant velocity vector $\mathbf{s} = (1.0, 2.0)^T$ by additive white Laplacian–Morgenstern bivariate noise $\mathbf{n}(k, l)$ with zero-mean vector, equal standard deviations

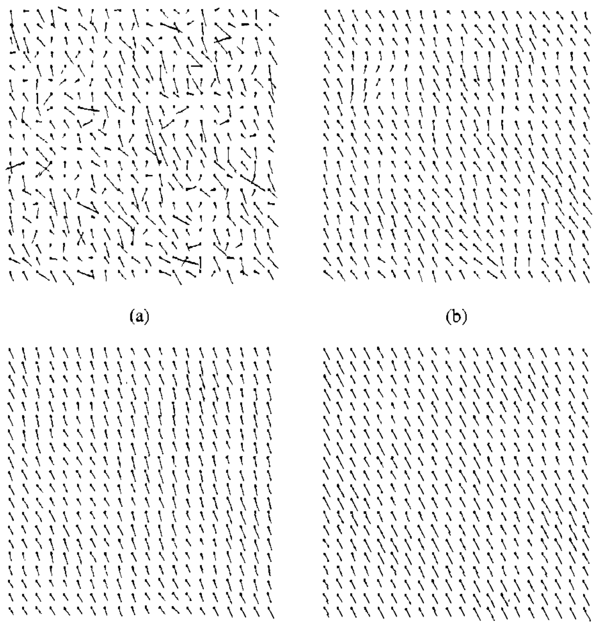


Fig. 2. Artificially generated velocity field produced by corrupting a constant velocity vector corrupted by additive white Laplacian–Morgenstern bivariate noise.

$\sigma_i = \sqrt{2}$, $i = 1, 2$ in both channels, and parameter $\alpha = 1.0$. The original vector field has been normalized appropriately in order to show the effect of noise. All the filtered outputs are shown normalized by the maximum value of each output channel. The recursive estimation of the constant velocity vector from the past filter outputs has been found quite successful. More specifically, the constant velocity vector $\mathbf{s}(k, l)$ has initially been estimated from the past filter outputs lying within a causal filter quadrant window of size 11×11, which has as right-most lower corner at the point where the constant velocity vector $\mathbf{s}(k, l)$ is to be estimated. The filtering is initialized by using the marginal median of the input vector-valued observations as an initial estimate of $\mathbf{s}$. The moments of the order statistics that are involved in the design procedure have been calculated based on estimates of the marginal and joint probability density functions of input data vectors. The same procedure has also been applied to the design of single-channel L filters. Fig. 2(b) shows the output of the marginal median filter of size 3×3. The filtered velocity field by two unbiased single-channel L filters of the same size that operate on each channel independently is shown in Fig. 2(c). Finally, Fig. 2(d) shows the result when the original data is filtered with the 3×3 unbiased two-channel marginal L filter. It is clearly seen that the 3×3 unbiased two-channel marginal L filter has superior performance. The first column in Table VI summarizes the noise reduction achieved at the output of several multichannel filters. For the unbiased single- and two-channel filters, the NR that would be achieved if the constant velocity $\mathbf{s}(k, l)$ and the distributional model were known is also listed in parentheses.

Subsequently, we consider a velocity field composed of two regions such that the velocity vectors are constant within each region but differ from the velocity vectors within the other region. In this case, the velocity field exhibits both horizontal and vertical transitions (i.e., edges). A pictorial description of the original uncorrupted vector field is given in Fig. 3. Such an artificially generated vector field is then corrupted by additive white Laplacian-Morgenstern bivariate noise $\mathbf{n}(k, l)$ with zero-mean vector, equal standard deviations $\sigma_i = \sqrt{2}$, $i = 1, 2$ in both channels, and parameter $\alpha = 1.0$. The portion of the vector field that is displayed in all figures is shown overlaid in Fig. 3. Fig. 4(a) displays the 20×20 block of the original noisy vector field. The noisy data vectors are shown normalized within each region by properly selecting convenient scaling factors in both channels in order to demonstrate the effect of noise. It can be seen that the noise is more prominent in Region 1, where the constant velocity $\mathbf{s}_1(k, l) = (1.0, 2.0)^T$ has been severely corrupted. In Region 2, the components of the constant velocity $\mathbf{s}_2(k, l) = (5.0, -10.0)^T$ are large enough compared with the noise standard deviation in each channel. Therefore, the velocity vector has been effected less by noise in Region 2. In the following, the components of all filter outputs are shown normalized by the maximum value of each channel within each region. Fig. 4(b) shows the result of vector median based on the L_1 norm (VM_{L_1}). The filter window is of size 3×3. Fig. 4(c) shows the result of the 3×3 arithmetic mean within each homogeneous region determined by the edge detection algorithm (65), that is, Fig. 4(b) shows the results of the 3×3 marginal median. In the sequel, (65) has been used to find the points that belong to a region where a horizontal edge occurs by employing windows W_N and W_S (Fig. 1(b)) and to find the points that belong to a segment where a vertical edge occurs by employing windows W_L and W_R (Fig. 1(c)). All these points form the transition region. On the other side of the transition region, two homogeneous regions result. Within each homogeneous region, two 3×3 single-channel L filters are used to filter the velocity components independently. Within the transition region, the 3×3 marginal median is applied. The output field is shown in Fig. 4(c). Finally, Fig. 4(d) shows the result when the original data are filtered by the 3×3

TABLE VI
NOISE REDUCTION (IN DECIBELS) FOR VARIOUS ESTIMATORS OF SIZE 3×3 IN THE CASE OF I) A CONSTANT VELOCITY FIELD AND II) A VELOCITY FIELD COMPOSED OF TWO REGIONS OF CONSTANT VELOCITY CORRUPTED BY ADDITIVE WHITE LAPLACIAN-MORGENSTERN BIVARIATE NOISE (FILTER LENGTH $N = 5$)

Filter	NR			
	Constant Field		Field with Edges	
$\mathcal{R}_E$ ($j = 3$)	-		-2.412	[10]
arithmetic mean (II: without segmentation information)	-8.001	[4]	-3.491	[9]
vector median L_2	-7.124	[7]	-5.426	[8]
multichannel MTM	-		-5.991	[7]
vector median L_1	-7.728	[6]	-6.292	[6]
generalized vector median	-7.754	[5]	-	
arithmetic mean (II: with segmentation information)	-		-6.630	[5]
multichannel DW-MTM (small window size 3×3)	-		-6.926	[4]
marginal median	-8.760	[3]	-7.056	[3]
unbiased single channel L-filter	-8.309 (-10.211)	[2]	-7.655	[2]
unbiased two-channel L-filter	-10.398 (-11.064)	[1]	-8.087	[1]

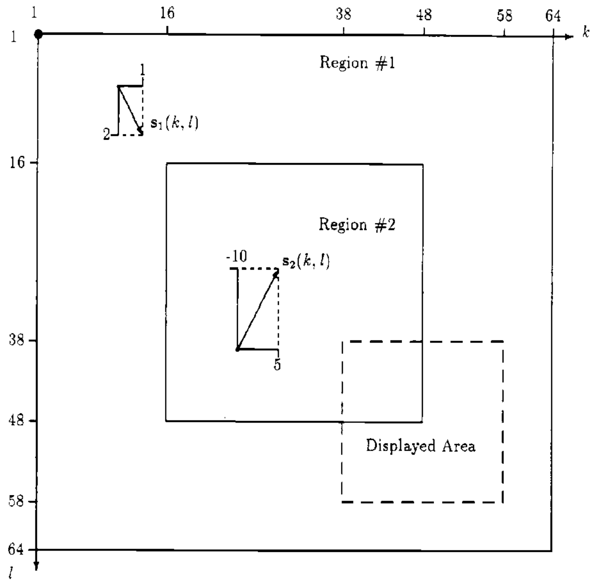


Fig. 3. Pictorial description of the uncorrupted vector field composed of two regions of constant velocity.

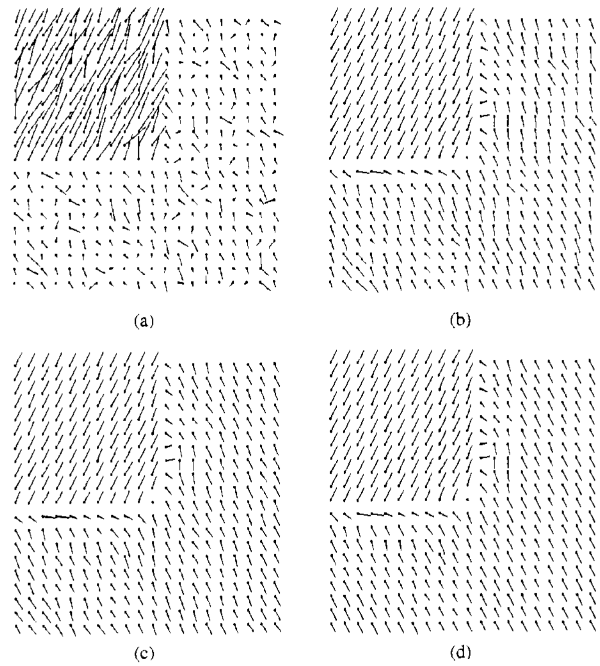


Fig. 4. Velocity field composed of two regions of constant velocity corrupted by additive white Laplacian–Morgenstern bivariate noise.

unbiased two-channel L filter within the homogeneous regions and the 3×3 marginal median in the transition region. Both the unbiased single-channel L filter as well as the unbiased two-channel L filter within each homogeneous region have been designed by using the estimation procedures developed above for calculating $\hat{\mathbf{R}}_2$ and $\hat{\underline{\mu}}_2$. A single estimate for $\mathbf{s}(k, l)$ has been found for each homogeneous region by employing the marginal median of the vector-valued observations within each homogeneous region. The second column in Table VI summarizes the noise reduction achieved at the filter output for several estimators. In conclusion, two-channel L filters can be used efficiently in the filtering of velocity fields produced by motion-estimation algorithms.

VI. Conclusion

The extension of single-channel L filters to the multichannel case has been discussed in this paper. The subordering principle of marginal data ordering has been used. The design of multichannel L filters based on marginal data ordering using the MSE as fidelity criterion has been presented. Both the unconstrained minimization of the MSE and the constrained minimization subject to the constraints of the unbiased and location-invariant estimation have been treated. A unified framework based on vector quantization for a discrete calculation of the required moments of the bivariate order statistics has been described. The derivation of a suitable long-tailed distribution, namely, the Laplacian distribution which belongs to the Morgenstern family and the design of a noise generator that obeys the above-mentioned distribution, has been presented. It has been shown by simulations that the unconstrained and unbiased two-channel marginal L filters attain a higher noise reduction index than other multichannel nonlinear filters (such as the vector median, the marginal α-trimmed mean, the marginal median, the multichannel modified trimmed mean, the multichannel double-window trimmed mean, and the multivariate ranked-order estimator $\mathcal{R}_E$) or their single-channel counterparts. Effective estimation procedures of the uncorrupted (true) multichannel signal $\mathbf{s}$ either for a multichannel constant or for a nonconstant signal corrupted by additive white multivariate noise have been developed. Furthermore, the design of multichannel marginal L filters based on estimates of the marginal and joint probability density functions of input vector-valued observations has been discussed. Applications to the filtering of artificially generated velocity fields have also been described.

References

[1] A. C. Bovik, T. S. Huang, and D. C. Munson, Jr., "A generalization of median filtering using linear combinations of order statistics," *IEEE Trans. Acoust. Speech Signal Processing*, vol. ASSP-31, no. 6, pp. 1342–1350, Dec. 1983.

[2] I. Pitas and A. N. Venetsanopoulos, *Nonlinear Digital Filters: Principles and Applications*. Norwell, MA: Kluwer, 1990.

[3] C. G. Boncelet, Jr., "Algorithms to compute order statistic distributions," *SIAM J. Sci. Stat. Comput.*, vol. 8, no. 5, pp. 868–876, Sept. 1987.

[4] F. Palmieri and C. G. Boncelet, Jr., "Optimal MSE linear combination of order statistics for restoration of Markov processes," in *Proc. 20th Ann. Conf. Inform. Sci. Syst.* (Princeton, NJ), Mar. 1986.

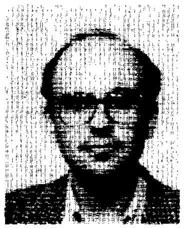
[5] L. Naaman and A. C. Bovik, "Least-squares order statistic filters for signal restoration," *IEEE Trans. Circuits Syst.*, vol. 38, no. 3, pp. 244–257, Mar. 1991.

[6] J. Astola, P. Haavisto, and Y. Neuvo, "Vector median filters," *Proc. IEEE* , vol. 78, no. 4, pp. 678–689, Apr. 1990.

[7] I. Pitas, "Marginal order statistics in color image filtering," *Opt. Eng.*, vol. 29, no. 5, pp. 495–503, May 1990.

[8] I. Pitas and P. Tsakalides, "Multivariate ordering in color image restoration," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 1, no. 3, pp. 247–259, Sept. 1991.

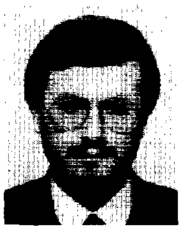
[9] R. C. Hardie and G. R. Arce, "Ranking in R^p and its use in multivariate image estimation," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 1, no. 2, pp. 197–209, June 1991.
[10] S. A. Kassam and M. Aburdene, "Multivariate median filters and their extensions," in *Proc. IEEE Int. Symp. Circuits Syst.*, June 1991, pp. 85–88.
[11] P. J. Huber, *Robust Statistics*. New York: Wiley, 1981.
[12] F. Hampel, E. Ronchetti, P. Rousseeuw, and W. Stahel, *Robust Statistics*. New York: Wiley, 1986.
[13] H. A. David, *Order Statistics*. New York: Wiley, 1981.
[14] V. Barnett, "The ordering of multivariate data," *J. R. Statist. Soc. A*, vol. 139, pt. 3, pp. 318–354, 1976.
[15] V. Barnett and T. Lewis, *Outliers in Statistical Data*. New York: Wiley, 1984.
[16] A. Papoulis, *Probability, Random Variables and Stochastic Processes*. New York: McGraw-Hill, 1984.
[17] R. Bellman, *Introduction to Matrix Analysis*. New York: McGraw-Hill, 1970, 2nd ed.
[18] D. G. Edelen and A. D. Kydoniefs, *An Introduction to Linear Algebra for Science and Engineering*. Amsterdam: Elsevier, 1972.
[19] A. C. Bovik and D. C. Munson, "Edge detection using median comparisons," *Comput. Vision Graphics Image Processing*, vol. 33, pp. 377–389, 1986.
[20] J. Max, "Quantizing for minimum distortion," *IRE Trans. Inform. Theory*, vol. IT-6, pp. 7–12, 1960.
[21] S. P. Lloyd, "Least squares quantization in PCM," *IEEE Trans. Inform. Theory*, pp. 127–135, Mar. 1982.
[22] R. J. Vaughan and W. N. Venables, "Permanent expressions for order statistic densities," *J. Roy. Stat. Soc.*, vol. 34, pp. 308–310, 1972.
[23] W. Maurer and B. H. Margolin, "The multivariate inclusion-exclusion formula and order statistics from dependent variates," *Annals Stat.*, vol. 4, no. 6, pp. 1190–1199, 1976.
[24] W. Feller, *An Introduction to Probability Theory and Its Applications*. New York: Wiley, 1968, vol. 1, 3rd ed.
[25] R. M. Meyer, "Note on a 'multivariate' form of Bonferroni's inequalities," *Ann. Math. Stat.*, vol. 40, no. 2, pp. 692–693, 1969.
[26] N. L. Johnson and S. Kotz, *Distributions in Statistics: Continuous Multivariate Distributions*. New York: Wiley, 1972.
[27] L. Devroye, *Non-Uniform Random Variate Generation*. New York: Springer-Verlag, 1986.
[28] J. Dagpunar, *Principles of Random Variate Generation*. Oxford: Clarendon, 1988.
[29] N. Nikolaidis, I. Pitas, and T. Kilindris, "Application of directional statistics to nonlinear vector field filtering and analysis," in *Proc. 1993 IEEE Winter Workshop Nonlinear Digital Signal Processing* (Tampere, Finland), January 1993, pp. 1.2-4.1–1.2-4.6.
[30] Y. H. Lee and S. A. Kassam, "Generalized median filtering and related nonlinear filtering techniques," *IEEE Trans. Acoust. Speech Signal Processing*, vol. ASSP-33, no. 3, pp. 672–683, June 1985.



Constantine Kotropoulos (M'93) was born in Kavala, Greece, in 1965. He received the Diploma degree with honors in electrical engineering in 1988 and the Ph.D. degree in electrical and computer engineering in 1993, both from the Aristotle University of Thessaloniki.

From 1989 to 1993, he was a research and a teaching assistant in the Department of Electrical and Computer Engineering at the same university, where he is currently a senior researcher. He has also conducted research in the Signal Processing Laboratory at Tampere University of Technology, Finland, during the summer of 1993. His current research interests include nonlinear digital signal processing, detection and estimation theory, pattern recognition, neural networks, and computer vision with applications to medical and geophysical images.

Dr. Kotropoulos has been a scholar of the State Scholarship Foundation of Greece and the Bodosaki Foundation. He is a member of the Technical Chamber of Greece.



Ioannis Pitas (SM'92) received the Diploma of electrical engineering in 1980 and the Ph.D. degree in electrical engineering in 1985, both from the University of Thessaloniki, Greece.

Since 1993, he has been an Associate Professor at the University of Thessaloniki. He has also been elected full Professor in the Department of Computer Science at the same University. From 1980 to 1993, he served as Scientific Assistant, Lecturer, and Assistant Professor at the same University. He served as a Visiting Research Associate at the University of Toronto, Canada, University of Erlangen-Nuernberg, Germany, Tampere University of Technology, Finland, and as Visiting Assistant Professor at the University of Toronto. He was lecturer in short courses for continuing education. His current interests are in the areas digital image processing, multidimensional signal processing, and computer vision. He has published over 113 papers and contributed to eight books in his area of interest. He is the co-author of the book *Nonlinear Digital Filters: Principles and Applications* (Boston: Kluwer, 1990). He is author of the book *Digital Image Processing Algorithms* (Englewood Cliffs: Prentice-Hall, 1993). He is editor of the book *Parallel Algorithms and Architectures for Digital Image Processing, Computer Vision and Neural Networks* (New York: Wiley, 1993).

Dr. Pitas has been member of the European Community ESPRIT Parallel Action Committee. He has also been invited speaker and/or member of the program committee of several scientific conferences and workshops. He is associate editor of IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS 6.