Nonlinear Ultrasonic Image Processing Based on Signal-Adaptive Filters and Self-Organizing Neural Networks

C. Kotropoulos, *Member, IEEE,* X. Magnisalis, I. Pitas, and M. G. Strintzis, *Senior Member, IEEE*

Abstract—**Two approaches for ultrasonic image processing are examined. First, signal-adaptive maximum likelihood (SAML) filters are proposed for ultrasonic speckle removal. It is shown that in the case of displayed ultrasound (US) image data the maximum likelihood (ML) estimator of the original (noiseless) signal closely resembles the L_2 mean which has been proven earlier to be the ML estimator of the original signal in US B-mode data. Thus, the design of signal-adaptive L_2 mean filters is treated for US B-mode data and displayed US image data as well. Secondly, the segmentation of ultrasonic images using self-organizing neural networks (NN) is investigated. A modification of the learning vector quantizer (L_2 LVQ) is proposed in such a way that the weight vectors of the output neurons correspond to the L_2 mean instead of the sample arithmetic mean of the input observations. The convergence in the mean and in the mean square of the proposed L_2 LVQ NN are studied. L_2 LVQ is combined with signal-adaptive filtering in order to allow preservation of image edges and details as well as maximum speckle reduction in homogeneous regions.**

I. Introduction

SPECKLE noise is a special kind of noise encountered in ultrasound (US) B-mode data as well as in images formed by laser beams or in synthetic aperture radar images [1], [2]. Suppression of speckle is desirable in order to enhance the quality of US images and therefore to increase the diagnostic potential of US examinations.

The statistical properties of speckle have been studied in [1]–[4]. The detection of focal lesions from the point of view of communication systems has been considered in [5]. Statistical nonlinear filters for speckle noise reduction on radar images have been examined in [6]. Bispectral restoration of speckle-degraded images performed on logarithmically transformed images has been proposed in [7] based on the assumption of a multiplicative noise model for speckle. Two approaches for speckle suppression in US B-mode data have been investigated in [8], [9]. The use of L_2 mean filter has been proposed based on the assumption that speckle is modeled as multiplicative noise distributed according to Rayleigh distribution. Optimal nonlinear filters based on linear combinations of the order statistics have also been designed for the estimation of a constant original (noiseless) signal corrupted by multiplicative Rayleigh noise. Homomorphic filtering could also be applied for speckle suppression [10].

A successful speckle suppression algorithm must reduce the noise efficiently by increasing the signal-to-noise ratio (SNR) without affecting true tissue information. It is evident that linear [11] and nonlinear [10] spatial filtering techniques cannot suppress the noise effectively due to their spatial-invariant nature and their weakness in discriminating between the foreground and the background areas in the image. The nonstationary nature of images in addition to the limitations of spatial filtering have motivated many researchers towards the investigation of filters which adjust their smoothing properties at each point of the image according to the local image content [12]–[19].

On the other hand, neural networks (NN) [20], [21] is a rapidly expanding research field which attracted the attention of scientists and engineers in the last decade. A large variety of artificial neural networks has been developed based on a multitude of learning techniques and having different topologies [21], [22]. They have applied successfully in many research fields. Neural networks have also found application for ultrasonic image analysis and diagnosis [23], [24].

Motivated by the fact that signal-adaptive filters fulfill the requirements for successful speckle suppression outlined above and the attention they have attracted in the literature, signal-adaptive maximum likelihood (SAML) filters for ultrasonic speckle suppression are proposed. The relative simplicity of the learning vector quantizer (LVQ), its ability to work in unsupervised mode in addition to its success in image segmentation problems, have led us to use a variant of the LVQ that is more suitable for ultrasonic image segmentation. Therefore, the main contribution of this paper is threefold: (1) The design of signal-adaptive maximum likelihood filters for ultrasonic speckle suppression both in the case of US B-mode data as well as in the case of displayed ultrasonic image data. (2) The derivation and study of convergence of a variant of LVQ neural network for ultrasonic image segmentation. (3) The design of a combined scheme where self-organizing NN are used in conjunction with signal-adaptive filters in order to allow image detail and edge preservation as well as maximum speckle reduction in homogeneous regions.

More specifically, the starting point is ultrasonic speckle modeling as multiplicative Rayleigh distributed noise (in the

Manuscript received July 8, 1992. The associate editor coordinating the review of this paper and approving it for publication was Dr. M. Ibrahim Sezan. This work has partially been supported by ESPRIT Basic Research Action project NAT (7130). C. Kotropoulos has also been supported by a scholarship from the State Scholarship Foundation of Greece and the Bodosaki Foundation.

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

IEEE Log Number 9213629.

case of US B-mode data), or as signal-dependent Gaussian noise (in the case of displayed US image data). The first model refers to envelope-detected US signal [2], [3], [25]. The second model describes more accurately ultrasonic images where the displayed image data have undergone excessive manipulation (e.g. logarithmic compression, low and high-pass filtering, postprocessing, etc.) [25]. It is proven that the maximum likelihood (ML) estimator of the original (noiseless) signal for the signal-dependent Gaussian speckle closely resembles the L_2 mean [26] which has already been proven to be the ML estimator of the original signal in US B-mode data [8], [9]. This fact has led us to design signal-adaptive maximum likelihood filters both for multiplicative Rayleigh speckle and signal-dependent speckle. Besides speckle filtering, the important problem of ultrasonic image segmentation by using self-organizing neural networks is also investigated. A modification of the learning vector quantizer (L_2 LVQ) is proposed so that the weight vectors of the output neurons correspond to the L_2 mean instead of the sample arithmetic mean of the input observations. The convergence in the mean and in the mean square of the proposed L_2 LVQ NN are studied. Further insight into the performance of the proposed L_2 LVQ is gained by an illustration of the convergence analysis for mutually independent 2-D Rayleigh distributed observations. The ability of the L_2 LVQ to segment ultrasonic images in various US image regions is combined with signal-adaptive filtering in order to allow image detail and edge preservation as well as maximum speckle reduction in homogeneous regions. Experimental results show that segmentation combined with filtering leads to higher speckle reduction without loss of image detail.

The outline of this paper follows. The design of signal-adaptive maximum likelihood filters is described in Section II. The derivation of L_2 LVQ algorithm, the study of its convergence and an illustration of the convergence analysis for mutually independent 2-D Rayleigh distributed observations are treated in Section III. Experimental results are included in Section IV and conclusions are drawn in Section V.

II. Signal-Adaptive Maximum Likelihood Filter Design

The design of signal-adaptive maximum likelihood filters for speckle suppression is examined in this section. First, an overview of the signal-adaptive filter model is provided. Secondly, two different models for speckle noise are considered, namely, the multiplicative noise model and the signal-dependent one. The ML estimator of the original (noiseless) signal for the signal-dependent noise model is derived. Subsequently, the derivation of the weighting factor that approximates the local SNR in the filter window is treated for both the above-mentioned models.

A. Signal-Adaptive Filter Model

An image x can be considered to consist of two parts: a low-frequency part x_L and a high-frequency part x_H, i.e., $x = x_L + x_H$ [10]. The low-frequency component is dominant in homogeneous image regions, whereas the high-frequency component is dominant near edges. The low-pass component can be estimated by a local estimator, e.g., the arithmetic mean or the median of the observations (i.e., pixel values) in a window $\mathcal{W}$ surrounding the current pixel (k,l) [10], [14]. The ML estimator $\hat{s}_{ML}(k,l)$ of the original signal $s(k,l)$ based on the observations $x(k-i,l-j) \in \mathcal{W}$ is proposed as the low-frequency component in this paper. Thus, the output of the signal-adaptive filter, i.e., the estimate of the original signal at (k,l), is given by

$$\hat{s}(k,l) = \hat{s}_{ML}(k,l) + \beta(k,l)[x(k,l) - \hat{s}_{ML}(k,l)] \quad (1)$$

where

$x(k,l)$ is the noisy observation at pixel (k,l)
$\hat{s}_{ML}(k,l)$ is the maximum likelihood estimate of $s(k,l)$ based on the observations $x(k-i,l-j) \in \mathcal{W}$
$\beta(k,l)$ is a weighting factor, approximating the local SNR over the window $\mathcal{W}$
$\hat{s}(k,l)$ is the signal-adaptive filter output at pixel (k,l).

When $\beta(k,l)$ approaches 1, the actual observation is preserved by the suppression of the low-pass component $\hat{s}_{ML}(k,l)$. When it is close to 0, maximum noise reduction is performed, since the high-frequency component is suppressed. In the remainder of this section, the ML estimator of $s(k,l)$ as well as the weighting factor $\beta(k,l)$ will be derived for the following two distinct noise models: 1) multiplicative Rayleigh speckle, and 2) signal-dependent Gaussian noise.

B. Derivation of the ML Estimator of the Original Signal

1) Multiplicative Noise Model: It has been shown [1]–[3] that the observed envelope-detected signal x can be considered as a Rayleigh random variable (r.v.) having probability density function

$$f_x(x) = \frac{x}{\sigma^2} \exp\left[-\frac{x^2}{2\sigma^2}\right], \quad x > 0. \quad (2)$$

Speckle can be modeled as

$$x = sn \quad (3)$$

where x is the observed envelope-detected signal, s is the original signal and n is a noise term statistically independent of s. It has been proven [8], [9] that the ML estimator of a constant original signal $s = S$ based on N observations $X_1, X_2, \cdots, X_N$ for model (3), assuming that n is a Rayleigh r.v. with unity expected value is given by

$$\hat{s}_{ML} = \frac{\sqrt{\pi}}{2} \sqrt{\frac{1}{N} \sum_{i=1}^{N} X_i^2} \quad (4)$$

which is equivalent to the L_2 mean [26] scaled by a factor of $\frac{\sqrt{\pi}}{2}$.

2) Signal-Dependent Noise Model: For displayed US images, a realistic image formation model is [25]

$$x = s + s^{1/2} n. \quad (5)$$

In the following, we assume that n is a zero-mean Gaussian random variable, i.e., its pdf is given by

$$f_n(n) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{n^2}{2\sigma^2}\right]. \quad (6)$$

The conditional density of an observation x assuming that $s = S, S > 0$ is given by [27]

$$f_{x|s}(X|S) = \frac{1}{\sigma\sqrt{2\pi S}} \exp\left[-\frac{(X-S)^2}{2\sigma^2 S}\right]. \quad (7)$$

Let us suppose that we have a set of N observations $x_i, i = 1, \cdots, N$ comprising the vector $\mathbf{x} = (x_1, \cdots, x_N)^T$, obtained from measurements in a window $\mathcal{W}$ around the pixel whose value s is to be estimated. The joint conditional pdf of the observations assuming $s = S$ is

$$f_{\mathbf{x}|s}(\mathbf{X}|S) = \prod_{i=1}^{N} f_{x_i|s}(X_i|S)$$
$$= \left(\frac{1}{\sigma\sqrt{2\pi S}}\right)^N \exp\left[-\frac{1}{2\sigma^2 S}\sum_{i=1}^{N}(X_i - S)^2\right]. \quad (8)$$

The ML estimate of $s = S$, $\hat{s}_{ML}$, maximizes the log-likelihood function $\ln f_{\mathbf{x}|s}(\mathbf{X}|S)$ [28]. It can be easily proven that $\hat{s}_{ML}$ is given by

$$\hat{s}_{ML} = -\frac{\sigma^2}{2} + \sqrt{\frac{\sigma^4}{4} + \frac{1}{N}\sum_{i=1}^{N} X_i^2}. \quad (9)$$

It is seen that (9) closely resembles the L_2 mean.

C. Derivation of the Signal-Dependent Weighting Factor

Let $e = \mathrm{E}_{\mathcal{W}}[(s - \hat{s})^2]$ be the local mean squared error (MSE) between the noiseless signal $s(k,l)$ and its estimate $\hat{s}(k,l)$ obtained by (1), at pixel (k,l), where $\mathrm{E}_{\mathcal{W}}[\cdot]$ denotes a local expectation operator. Explicit reference to (k,l) can be suppressed for notation simplicity. We shall evaluate β at each pixel (k,l) of the image by minimizing e over a window $\mathcal{W}$ surrounding pixel (k,l)

$$e = \mathrm{E}_{\mathcal{W}}[(s - \hat{s}_{ML})^2] + \beta^2 \mathrm{E}_{\mathcal{W}}[(x - \hat{s}_{ML})^2]$$
$$-2\beta \mathrm{E}_{\mathcal{W}}[(s - \hat{s}_{ML})(x - \hat{s}_{ML})]. \quad (10)$$

Differentiating with respect to β and setting the result equal to zero we get

$$\beta = \frac{\mathrm{E}_{\mathcal{W}}[(s - \hat{s}_{ML})(x - \hat{s}_{ML})]}{\mathrm{E}_{\mathcal{W}}[(x - \hat{s}_{ML})^2]}. \quad (11)$$

In the sequel, $\mathrm{E}_{\mathcal{W}}[(s - \hat{s}_{ML})(x - \hat{s}_{ML})]$ will be evaluated for the two speckle models under study.

1) Multiplicative Noise Model: The numerator in (11) can be rewritten as

$$\mathrm{E}_{\mathcal{W}}[(s - \hat{s}_{ML})(x - \hat{s}_{ML})] = \mathrm{E}_{\mathcal{W}}[s(x - \hat{s}_{ML})]$$
$$- \hat{s}_{ML}E_{\mathcal{W}}[x - \hat{s}_{ML}] = \mathrm{E}_{\mathcal{W}}[s^2]$$
$$- 2\hat{s}_{ML}\mathrm{E}_{\mathcal{W}}[x] + \hat{s}_{ML}^2. \quad (12)$$

where we assume that $\mathrm{E}[n] = 1$ and s, n are independent, thus, $\mathrm{E}_{\mathcal{W}}[x] = \mathrm{E}_{\mathcal{W}}[s]$. The statistical independence of s and n implies that

$$\mathrm{E}[s^2] = \frac{\mathrm{E}[x^2]}{\mathrm{E}[n^2]} = \frac{\mathrm{E}[x^2]}{\mathrm{E}^2[n] + \mathrm{var}\,[n]} = \frac{E[x^2]}{1 + \frac{4-\pi}{\pi}}. \quad (13)$$

where $\frac{4-\pi}{\pi}$ is the variance of a Rayleigh r.v. having unity expected value. Therefore, (11) takes the form

$$\beta = 1 - \left(\frac{4-\pi}{4}\right)\frac{\mathrm{E}_{\mathcal{W}}[x^2]}{\mathrm{E}_{\mathcal{W}}[(x - \hat{s}_{ML})^2]}. \quad (14)$$

2) Signal-Dependent Noise Model: If n is zero-mean r.v. independent of s, it is obvious that $\mathrm{E}_{\mathcal{W}}[x] = \mathrm{E}_{\mathcal{W}}[s]$. Therefore

$$\mathrm{E}_{\mathcal{W}}[s(x - \hat{s}_{ML})] = \mathrm{E}_{\mathcal{W}}[s(s + s^{1/2}n - \hat{s}_{ML})]$$
$$= \mathrm{E}_{\mathcal{W}}[x^2] - \mathrm{E}_{\mathcal{W}}[x]\sigma^2 - \hat{s}_{ML}E_{\mathcal{W}}[x] \quad (15)$$

because

$$\mathrm{E}_{\mathcal{W}}[x^2] = \mathrm{E}_{\mathcal{W}}[(s + s^{1/2}n)^2] = \mathrm{E}_{\mathcal{W}}[s^2] + \mathrm{E}_{\mathcal{W}}[s]\sigma^2 \quad (16)$$

where σ^2 denotes the variance of noise. Substituting (15) into (11), the following signal-dependent weighting factor results

$$\beta = 1 - \frac{\mathrm{E}_{\mathcal{W}}[x]\sigma^2}{\mathrm{E}_{\mathcal{W}}[(x - \hat{s}_{ML})^2]}. \quad (17)$$

In general, β can be considered as a local signal-to-noise ratio measure [14]. The role of the signal-dependent weighting factor β (14) or (17) is to adjust the magnitude of the filtering to be performed by (1) as well as to adjust the size of window $\mathcal{W}$. We can start filtering by using initially a window of size 5×5 or 7×7. If the factor $\beta(k,l)$ becomes greater than an appropriate threshold β_t, the window size is decreased until the coefficient becomes less than the threshold or until the window reaches the size 3×3. Otherwise, the window is increased to its maximum size.

Subsequently, a modification of the signal-adaptive maximum likelihood filter that utilizes segmentation information obtained prior to the filtering process is discussed. An image can be segmented to regions representing various image characteristics by a variety of techniques that can be found in digital image processing literature [11] as well as by using neural networks [21], [22]. Let us denote by $\mathcal{L}(k,l)$ the class label assigned to pixel (k,l) of the original image by using an image segmentation algorithm. We shall assume that the original image is segmented into p classes. If the label assigned to the first class is 0, $\mathcal{L}(k,l)$ will vary between 0 and $p-1$. In the case of a two-class US image segmentation, the label 0 can be assigned to pixels belonging to the background and the label 1 can be assigned to pixels that belong to areas that have high signal activity. In the general case of a p-class US image segmentation, label 0 is assigned to background pixels and the remaining classes are labeled according to their signal activity (e.g. frequency content). The classes that have the higher signal activity correspond to image details (e.g. blood vessel boundaries). The more details an image region has the higher label is assigned to it. Let us also define by

$$\beta_{\mathrm{seg}}(k,l) = \frac{1}{(p-1)N(\mathcal{W})}\sum_{(k-i,l-j)\in\mathcal{W}} \mathcal{L}(k-i, l-j) \quad (18)$$

a quantity that represents the average class label that appears in the filter window $\mathcal{W}$ scaled in the range [0, 1]. In (18), $N(\mathcal{W})$ denotes the number of pixels included in the filter window (i.e., the window size). It can be seen that β_{seg} varies between 0 and 1. It tends to 0, when the pixels that belong to

the lowest signal activity class dominate in the filter window $\mathcal{W}$ surrounding pixel (k,l). On the contrary, it tends to 1 when the pixels in the filter window $\mathcal{W}$ surrounding pixel (k,l) that belong to the highest signal activity class are in the majority. Since $\beta_{\text{seg}} \in [0,1]$, it can be used as weighting factor in the signal-adaptive filter (1). It is seen that by using (18) as weighting factor, the classes that correspond to image details are preserved. Furthermore, if the signal-adaptive filter has to share the benefits of the signal-dependent weighting factor (14) or (17) as local SNR measure as well as of the factor (18) as local texture indicator, the following modified weighting factor $\hat{\beta}(k,l)$ is proposed

$$\hat{\beta}(k,l) = (1-\eta)\beta(k,l) + \eta\beta_{\text{seg}}(k,l) \tag{19}$$

where $0 \leq \eta \leq 1$. By selecting $\eta = 0$, then $\hat{\beta}(k,l) = \beta(k,l)$ and the segmentation information provided by the image segmentation algorithm is discarded. If $\eta = 1$, $\hat{\beta}(k,l) = \beta_{\text{seg}}(k,l)$ and the local SNR information is ignored. Otherwise, both information sources are taken into account. In order to design a signal-adaptive filter that uses the signal-dependent weighting factor (19), we have to segment the ultrasonic image first. In the next section, the segmentation of US images by using a variant of learning vector quantizer that is based on the L_2 mean is examined.

III. A Variant of LVQ Based on the L_2 Mean for Segmentation of Ultrasonic Images

LVQ is an autoassociative nearest-neighbor classifier which classifies arbitrary patterns into $p-$many classes using an error correction encoding procedure related to competitive learning [20], [21]. Let us assume a sequence of vector-valued observations $\mathbf{x}(t) \in \mathcal{R}^N$ and a set of variable reference vectors (i.e., output neurons) $\{\mathbf{w}_i(t); \mathbf{w}_i \in \mathcal{R}^N, i = 1,2,\cdots,p\}$. Let $\mathbf{w}_i(0)$ be randomly initialized. Competitive learning tries to find the best-matching reference vector $\mathbf{w}_c(t)$ to $\mathbf{x}(t)$. This vector is updated. After a large number of iterations, the different reference vectors tend to become specifically "tuned" to different domains of the input variable $\mathbf{x}$. LVQ belongs to the class of self-organizing neural networks where the reference vectors are ordered spatially. A neighborhood set $\mathcal{N}_c$ is defined around each winner $\mathbf{w}_c$. At each learning step, all the reference vectors within $\mathcal{N}_c$ are updated, whereas reference vectors outside $\mathcal{N}_c$ are left intact. When the learning procedure is led to equilibrium, it results in a partition of the domain of input vector-valued observations into regions called Voronoi neighborhoods. The reference vector of each Voronoi neighborhood is the centroid, e.g., the sample arithmetic mean of all input vectors belonging to that neighborhood.

Self-organizing neural networks, like LVQ and its variants (as the one discussed in this paper), constitute an attractive alternative to neural network architectures that involve hidden layers and are trained by using the backpropagation algorithm. One appealing characteristic of LVQ is its simplicity. It is a NN consisted of a single layer modifiable connections. For example, the learning procedure of LVQ is much simpler than backpropagation algorithm. In addition, its learning procedure is much faster than that of three layer networks that have the same total number of neurons (hidden neurons plus output ones) with the number of output neurons of LVQ, because the time-consuming training of hidden neurons is omitted. Furthermore, the important problem of selecting the number of hidden neurons that poses an additional difficulty is not faced in the case of LVQ network. Another attractive feature of LVQ is that it can be used both for unsupervised as well as for supervised learning. This is not the case for networks with hidden layers. The later implement only supervised learning, i.e., they require the availability of a reference signal that is used as a teacher.

In Section II it has been proven that the maximum likelihood estimator of the original noiseless image is the L_2 mean both in the case of pure multiplicative noise (3) as well as in the case of signal-dependent Gaussian noise (5). This observation motivated us to modify the standard LVQ algorithm so that the reference vectors correspond to the L_2 mean instead of the sample arithmetic mean. In the sequel, the proposed L_2 LVQ will be described, its convergence properties will be studied and an illustration of the convergence analysis for Rayleigh distributed input vectors will be treated.

A. Derivation of the L_2 LVQ Algorithm

Let us denote by $\mathbf{w}_i'$ the $(N \times 1)$ vector having as elements the weights comprising the reference vector $\mathbf{w}_i$ squared, i.e., $\mathbf{w}_i' = (w_{i1}^2, w_{i2}^2, \cdots, w_{iN}^2)^T$. Let also $\mathbf{x}'$ denote the following vector $\mathbf{x}' = (x_1^2, x_2^2, \cdots, x_N^2)^T$. Our goal is to place $\mathbf{w}_i'$ into the input space $\mathcal{R}^N$ in such a way that they minimize the mean squared value of a reconstruction error of the form

$$\varepsilon = \int_{\mathcal{X}} \|\mathbf{x}' - \mathbf{w}_c'\|^2 f(\mathbf{x}) d\mathbf{x} \tag{20}$$

where $\mathcal{X} \subseteq \mathcal{R}^N$ is the domain of the vector-valued observations $\mathbf{x}$, $d\mathbf{x}$ is the volume differential in the $\mathcal{R}^N$ space and $\mathbf{w}_c'$ is the winner vector. The winner vector is determined by minimizing

$$\|\mathbf{x}' - \mathbf{w}_c'\| = \min_i \{\|\mathbf{x}' - \mathbf{w}_i'\|\} \tag{21}$$

where $\|\cdot\|$ stands for the Euclidean distance metric. If the stochastic-gradient-descent algorithm [30] is applied to the minimization of ε in the $\mathbf{w}_c'$ space and the vectors of square weights are updated as blocks concentrated around the winner, the following recursive relations for updating the vectors of square weights yield

$$\begin{aligned} \mathbf{w}_i'(t+1) &= \mathbf{w}_i'(t) + \alpha(t)[\mathbf{x}'(t) - \mathbf{w}_i'(t)] \ \forall i \in \mathcal{N}_c(t) \\ \mathbf{w}_i'(t+1) &= \mathbf{w}_i'(t) \ \forall i \notin \mathcal{N}_c(t) \end{aligned} \tag{22}$$

where $\alpha(t)$ is a variable adaptation step defined as [21, 29]

$$\alpha(t) = 0.2\left[1 - \frac{t}{10000}\right]. \tag{23}$$

The updating (22) implements the unsupervised learning of the L_2 LVQ neural network. It can be seen that (22) resembles the following simple recursive formula for the computation of the

L_2 mean of $(N+1)$ observations based on the L_2 mean of the past N observations and the current one

$$w_{L2}^2(N+1) = w_{L2}^2(N) + \frac{1}{N+1}(x^2(N+1) - w_{L2}^2(N)). \quad (24)$$

The *recall procedure* of the L_2 LVQ is used to determine the class $\mathcal{C}_g$ represented by $\overline{\mathbf{w}_g'}$ to which the vector of squared input observations is most closely associated with, i.e.

$$\mathbf{x}(t) \in \mathcal{C}_g \text{ if } \|\mathbf{x}' - \overline{\mathbf{w}_g'}\| = \min_i^p \{\|\mathbf{x}' - \overline{\mathbf{w}_i'}\|\} \quad (25)$$

where $\overline{\mathbf{w}_i'}$ denotes the vector of squared weights of the i-th reference vector after the convergence of the learning procedure.

Having described the L_2 LVQ neural network, we proceed to the study of its convergence. Our aim is to derive the expected stationary state of the network, to find bounds on the rate of convergence and conditions on the adaptation step as well.

B. Convergence Analysis of L_2 LVQ Neural Network

The learning procedure of the LVQ has been represented by a Markov process whose states are the weight vectors $\mathbf{w}_i$ [31]. The Fokker-Planck differential equation describing the learning process in the vicinity of the equilibrium in terms of the distribution of weight-error vector has been derived. The average weight-error vector and the weight-error correlation matrix have been determined as well. A necessary and sufficient condition which guarantees convergence of the learning procedure to an asymptotic equilibrium map has also been proposed. The analysis presented in this paper extends the work reported in [31].

First, let us define the two types of convergence to be examined in the sequel. L_2 LVQ network *converges in the mean*, if the average vector of squared weights converges to the expected stationary state of the network as t approaches infinity. L_2 LVQ network *converges in the mean square*, if the trace of the correlation matrix of the squared weight-error vectors tends to zero or remains bounded as t approaches infinity. Both convergence in the mean and in the mean square will be studied for L_2 LVQ. The case of a constant adaptation step $\alpha(t) = \alpha$ will be considered for mathematical simplicity. Generalization for the optimal adaptation step sequence $\alpha(t) = 1/t$ [31] will also be considered. We shall confine ourselves to the analysis of a single-winner L_2 LVQ network, i.e., $\mathcal{N}_c(t) = \{c\}$. Our objective is to derive bounds on the overall time constant for any squared weight and on the trace of the correlation matrix of the squared weight-error vectors.

Let $\mathcal{V}_i(\mathbf{W}')$ denote the Voronoi neighborhood of the $i-$th output neuron with respect to the distance metric (21), i.e.

$$\mathcal{V}_i(\mathbf{W}') = \{\mathbf{x} \in \mathcal{X} \subseteq \mathcal{R}^N | \ \|\mathbf{x}' - \mathbf{w}_i'\| \leq \|\mathbf{x}' - \mathbf{w}_l'\| \quad l = 1, \cdots, p, l \neq i\} \quad (26)$$

where $\mathbf{W}' = \left(\mathbf{w}'^T_1 | \mathbf{w}'^T_2 | \cdots | \mathbf{w}'^T_p\right)^T$. The expected stationary state of the network is given by [31]

$$\overline{\mathbf{w}_i'} = \mathrm{E}[\mathbf{w}_i'] = \frac{\int_{\mathcal{V}_i(\overline{\mathbf{W}'})} \mathbf{x}' f(\mathbf{x}) d\mathbf{x}}{\int_{\mathcal{V}_i(\overline{\mathbf{W}'})} f(\mathbf{x}) d\mathbf{x}} \qquad i = 1, \cdots, p. \quad (27)$$

It is seen that (27) gives an implicit definition of the stationary state of L_2 LVQ. In the sequel, it will be assumed that $\overline{\mathbf{w}_i'}$ is known and our attention will be focused on the study of the rate of convergence to the stationary state.

Let $\mathbf{u}_i(t) = \mathbf{w}_i' - \overline{\mathbf{w}_i'}$ denote the $(N \times 1)$ vector of squared weight errors at time instant t. The average squared weight-error vector are given by $\mathrm{E}[\mathbf{U}(t)] = \mathbf{Y}(t)\mathrm{E}[\mathbf{U}(0)]$ with $\mathbf{U}(t) = \left(\mathbf{u}_1^T(t) | \cdots | \mathbf{u}_p^T(t)\right)^T$, where the expectation is with respect to the distribution of the deviations of the squared weights from the stationary state [31], [32]. $\mathbf{Y}(t)$ is the following $(Np \times Np)$ matrix

$$\mathbf{Y}(t) = \exp(-\mathbf{B}\int_0^t \alpha(\zeta) d\zeta) \quad (28)$$

where $\mathbf{B}$ is a $(Np \times Np)$ coefficient matrix which can be partitioned as follows

$$\mathbf{B} = \begin{bmatrix} \mathbf{B}_{11} & \mathbf{B}_{12} & \cdots & \mathbf{B}_{1p} \\ \mathbf{B}_{21} & \mathbf{B}_{22} & \cdots & \mathbf{B}_{2p} \\ \vdots & & \ddots & \vdots \\ \mathbf{B}_{p1} & \mathbf{B}_{p2} & \cdots & \mathbf{B}_{pp} \end{bmatrix}. \quad (29)$$

Each $\mathbf{B}_{kl}$ $k, l = 1, \cdots, p$ is a $(N \times N)$ square submatrix with mn-element given by

$$[\mathbf{B}_{kl}(\overline{\mathbf{W}'})]_{mn} = \left[w_{km}^2 \frac{\partial}{\partial w_{ln}^2} \hat{F}_k(\mathbf{W}') + \hat{F}_k(\mathbf{W}')\delta(k-l, m-n) - \frac{\partial}{\partial w_{ln}^2} \int_{\mathcal{V}_k(\mathbf{W}')} x_m^2 f(\mathbf{x}) d\mathbf{x} \right]_{\mathbf{W}' = \overline{\mathbf{W}'}} \quad (30)$$

where

$$\hat{F}_k(\mathbf{w}') = \int_{\mathcal{V}_k(\mathbf{W}')} f(\mathbf{x}) d\mathbf{x} \quad (31)$$

and $\delta(k-l, m-n)$ is the 2-D Kronecker-delta function. Let us assume that (the real) matrix $\mathbf{B}$ is symmetric. It must be noted that there is no such a guarantee in the general case of a process described by a multivariate linear Fokker-Planck equation [32]. Hopefully, the assumption that matrix $\mathbf{B}$ is symmetric is valid in our case (i.e., the coefficient matrix $\mathbf{B}$ defined in (29)–(31) and evaluated for the contaminated Rayleigh distribution) [39]. If $\mathbf{B}$ is symmetric, it is diagonalizable and possesses real eigenvalues [34]. Let $\lambda_i, i = 1, \cdots, q, q \leq Np$ be the distinct eigenvalues of matrix $\mathbf{B}$ and ρ_i be the degree of multiplicity of the eigenvalue λ_i where

$$\sum_{i=1}^{q} \rho_i = Np \quad (32)$$

$\mathbf{Y}(t)$ defined in (28) is a $(Np \times Np)$ matrix which can be evaluated as follows

$$\mathbf{Y}(t) = \sum_{i=1}^{Np} y_i(t) \mathbf{B}^{i-1} \quad (33)$$

where $\mathbf{B}^0 = \mathbf{I}$ is the $(Np \times Np)$ identity matrix and $y_i(t), i = 1, \cdots, Np$ are scalar functions to be determined. By applying

Caley-Hamilton's theorem we obtain [35]

$$\sum_{i=1}^{Np} y_i(t)\lambda_j^{i-1} = \exp(-\lambda_j \int_0^t \alpha(\zeta)d\zeta) \qquad j = 1, \cdots, q. \tag{34}$$

In the special case that matrix $\mathbf{B}$ possesses Np distinct eigenvalues, the application of Sylvester's theorem yields the following equation for $\mathbf{Y}(t)$ [35]

$$\mathbf{Y}(t) = \sum_{i=1}^{Np} \left(\prod_{j=1 j\neq i}^{Np} \frac{\mathbf{B} - \lambda_j \mathbf{I}}{\lambda_i - \lambda_j} \right) \exp\left(-\lambda_i \int_0^t \alpha(\zeta)d\zeta\right). \tag{35}$$

In the following, we shall assume that the eigenvalues of matrix $\mathbf{B}$ are distinct. Therefore, we shall confine ourselves to (35). For a constant adaptation step $\alpha(t) = \alpha$, (35) is rewritten as follows

$$\mathbf{Y}(t) = \sum_{i=1}^{Np} \left(\prod_{j=1 j\neq i}^{Np} \frac{\mathbf{B} - \lambda_j \mathbf{I}}{\lambda_i - \lambda_j} \right) \exp(-\lambda_i \alpha t) \tag{36}$$

whereas in the general case of a variable adaptation step $\alpha(t) = 1/t$

$$\begin{aligned} \mathbf{Y}(t) &= \lim_{t_0 \to 0} \sum_{i=1}^{Np} \left(\prod_{j=1 j\neq i}^{Np} \frac{\mathbf{B} - \lambda_j \mathbf{I}}{\lambda_i - \lambda_j} \right) \exp(-\lambda_i(\ln t - \ln t_0)) \\ &= \sum_{i=1}^{Np} \left(\prod_{j=1 j\neq i}^{Np} \frac{\mathbf{B} - \lambda_j \mathbf{I}}{\lambda_i - \lambda_j} \right) \frac{1}{t^{\lambda_i}} \end{aligned} \tag{37}$$

provided that $\lambda_i > 0, i = 1, \cdots, Np$.

By combining the definition of $\mathrm{E}[\mathbf{U}(t)]$ and (36)–(37), it is seen that a necessary and sufficient condition for the convergence in the mean is matrix $\mathbf{B}$ to be positive definite, i.e., $\lambda_i > 0$. The convergence is negative exponential in the case of a constant adaptation step. The convergence is hyperbolic when $\alpha(t) = 1/t$. In the former case, α cannot be bounded by the analysis made thus far.

For a constant adaptation step α, any squared weight converges in the mean to the stationary state (27) as a weighted sum of negative exponentials of the form $\exp(-\lambda_i \alpha t)$, because $\mathbf{Y}(t)$ is a square matrix of negative exponentials. The time τ_i required for each term to reach to $1/e$ of its initial value is given by $\tau_i = 1/(\alpha\lambda_i)$. By using the same reasoning as in the adaptive filter literature [36, 37], the overall time constant τ_a for any average squared weight can be bounded as follows

$$\frac{1}{\alpha\lambda_{\max}} \leq \tau_a \leq \frac{1}{\alpha\lambda_{\min}} \tag{38}$$

where $\lambda_{\min}$ and $\lambda_{\max}$ denote the smallest and largest eigenvalue of matrix $\mathbf{B}$.

The study of the convergence in the mean square will be focused on the case of a constant adaptation step for mathematical tractability. Let $\mathbf{C}(t)$ denote the correlation matrix of the squared weight-error vectors. $\mathbf{C}(t)$ is of dimensions $(Np \times Np)$ and has also the structure of (29) where the $(N \times N)$ square submatrices $\mathbf{C}_{kl}(t), k, l = 1, \cdots, p$ are defined by

$$\mathbf{C}_{kl}(t) = \mathrm{E}[\mathbf{u}_k(t)\mathbf{u}_l^T(t)]. \tag{39}$$

$\mathbf{C}(t)$ can be evaluated as follows [32]

$$\mathbf{C}(t) = \mathbf{Y}(t)\left[\mathbf{C}(0) + \alpha^2 \int_0^t \mathbf{Y}(\zeta)^{-1}\mathbf{D}(\mathbf{Y}(\zeta)^{-1})^T d\zeta\right]\mathbf{Y}(t)^T \tag{40}$$

where $\mathbf{C}(0)$ is the initial correlation matrix and $\mathbf{D}$ is a $(Np \times Np)$ matrix having the structure (29) with only the diagonal submatrices being non-zero, i.e., $\mathbf{D}_{kl} = \mathbf{0}_{N\times N}, k \neq l$. The mn-element of $\mathbf{D}_{kk}$ is given by

$$\begin{aligned} [\mathbf{D}_{kk}(\overline{\mathbf{W}'})]_{mn} = \Big[& w_{km}^2 w_{kn}^2 \hat{F}_k(\mathbf{W}') - w_{km}^2 \\ & \int_{\mathcal{V}_k(\mathbf{W}')} x_n^2 f(\mathbf{x})d\mathbf{x} - w_{kn}^2 \int_{\mathcal{V}_k(\mathbf{W}')} x_m^2 f(\mathbf{x})d\mathbf{x} \\ & + \int_{\mathcal{V}_k(\mathbf{W}')} x_m^2 x_n^2 f(\mathbf{x})d\mathbf{x} \Big]_{\mathbf{W}'=\overline{\mathbf{W}'}}. \end{aligned} \tag{41}$$

It can be seen that coefficient matrix $\mathbf{D}$ is symmetric as a result of its definition (41). Furthermore, it can be proven that matrix $\mathbf{D}$ is positive semidefinite in any case [33]. It is known [32] that the following equation holds for the time-derivative $\dot{\mathbf{C}}(t)$ of the correlation matrix $\mathbf{C}(t)$

$$\dot{\mathbf{C}}(t) = -\alpha[\mathbf{B}\mathbf{C}(t) + \mathbf{C}(t)\mathbf{B}^T] + \alpha^2\mathbf{D}. \tag{42}$$

Let $J(t)$ denote the trace of the correlation matrix $\mathbf{C}(t)$

$$J(t) = \mathrm{tr}[\mathbf{C}(t)] = \mathrm{E}\left[\sum_{i=1}^{p} \mathbf{u}_i^T(t)\mathbf{u}_i(t)\right]. \tag{43}$$

It can be shown that $J(t)$ is bounded as follows

$$\begin{aligned} \left(J(0) - \frac{\alpha\mathrm{tr}[\mathbf{D}]}{2\lambda_{\max}}\right)\exp(-2\alpha\lambda_{\max}t) + \frac{\alpha\mathrm{tr}[\mathbf{D}]}{2\lambda_{\max}} \leq J(t) \\ \leq \left(J(0) - \frac{\alpha\mathrm{tr}[\mathbf{D}]}{2\lambda_{\min}}\right)\exp(-2\alpha\lambda_{\min}t) + \frac{\alpha\mathrm{tr}[\mathbf{D}]}{2\lambda_{\min}} \end{aligned} \tag{44}$$

where $J(0) = \mathrm{tr}[\mathbf{C}(0)]$. The derivation of inequalities (44) is given in the appendix. It is seen that if the adaptation step is constant, we can only require $J(t)$ to remain bounded. Let J_b denote the maximum allowed deviation of $J(t)$ from zero when the exponential factor in the upper bound of (44) has practically converged to zero, i.e., at $t = 4\tau'_{\max}$. A sufficient condition for $J(t)$ to remain bounded is given by

$$0 < \alpha < \frac{2\lambda_{\min}J_b}{\mathrm{tr}[\mathbf{D}]}. \tag{45}$$

In the following, an illustration of the convergence analysis will be treated.

C. Illustration of the Convergence Analysis for Mutually Independent 2-D Rayleigh Distributed Observations

We shall confine ourselves to a network having two neurons in both layers for mathematical tractability. Such an approach leads to analytical results expressed in closed formulae and allows an easy interpretation. Let $\mathbf{w}'_1 = (w_{11}^2, w_{12}^2)^T$ and $\mathbf{w}'_2 = (w_{21}^2, w_{22}^2)^T$ be the vectors of squared weights for

the two output neurons. 2-D input observations will be mutually independent and will obey the contaminated Rayleigh distribution

$$f(x_1,x_2)=(1-\epsilon)\frac{x_1x_2}{\sigma_{11}^2\sigma_{12}^2}\exp\left[-\left(\frac{x_1^2}{2\sigma_{11}^2}+\frac{x_2^2}{2\sigma_{12}^2}\right)\right]+\epsilon\frac{x_1x_2}{\sigma_{21}^2\sigma_{22}^2}\exp\left[-\left(\frac{x_1^2}{2\sigma_{21}^2}+\frac{x_2^2}{2\sigma_{22}^2}\right)\right]. \quad (46)$$

The independence assumption is used only to facilitate the statistical description of 2-D input observations having Rayleigh marginal distributions. The contaminated Rayleigh distribution has been chosen because it is the simplest way to produce two clusters in the 2-D Euclidean space $\mathcal{R}^2$. The definition of the Voronoi neighborhoods of the output neurons (26) leads to the following equations

$$\mathcal{V}_1(\mathbf{W}')=\left\{(x_1,x_2)\in\mathcal{X}\subseteq\mathcal{R}^2 \mid (w_{11}^2-w_{21}^2)x_1^2+(w_{12}^2-w_{22}^2)x_2^2\geq\frac{(w_{11}^4+w_{12}^4)-(w_{21}^4+w_{22}^4)}{2}\right\}$$
$$\mathcal{V}_2(\mathbf{W}')=\mathcal{X}-\mathcal{V}_1(\mathbf{W}'). \quad (47)$$

Let $d(\mathbf{W}')=a(\mathbf{W}')-b(\mathbf{W}')x_1^2$, where

$$a(\mathbf{W}')=\frac{(w_{11}^4+w_{12}^4)-(w_{21}^4+w_{22}^4)}{2(w_{12}^2-w_{22}^2)}$$
$$b(\mathbf{W}')=\frac{w_{11}^2-w_{21}^2}{w_{12}^2-w_{22}^2}. \quad (48)$$

Depending on the sign of the quantities $w_{11}^2-w_{21}^2$, $w_{12}^2-w_{22}^2$ and $a(\mathbf{W}')$ all possible configurations for the Voronoi neighborhoods $\mathcal{V}_1(\mathbf{W}')$ and $\mathcal{V}_2(\mathbf{W}')$ are listed in Table I. The upper inequality symbol refers to $\mathcal{V}_1(\mathbf{W}')$ whereas the lower one to $\mathcal{V}_2(\mathbf{W}')$. In the following, we shall assume that

$$w_{11}^2>w_{21}^2 \text{ and } w_{12}^2<w_{22}^2. \quad (49)$$

Inequalities (49) guarantee good separation between the two clusters. Let $\epsilon_1=1-\epsilon$ and $\epsilon_2=\epsilon$. For the contaminated Rayleigh distribution (46) we have

$$\hat{F}_1(\mathbf{W}')=\sum_{k=1}^{2}\epsilon_k\left(1-\exp\left[-\frac{a(\mathbf{W}')}{2\sigma_{k2}^2}\right]\frac{\sigma_{k2}^2}{\sigma_{k2}^2-b(\mathbf{W}')\sigma_{k1}^2}\right)$$
$$\hat{F}_2(\mathbf{W}')=1-\hat{F}_1(\mathbf{W}') \quad (50)$$

and

$$\int_{\mathcal{V}_1(\mathbf{W}')}x_1^2f(x_1,x_2)dx_1dx_2=2\sum_{k=1}^{2}\epsilon_k\sigma_{k1}^2\left(1-\exp\left[-\frac{a(\mathbf{W}')}{2\sigma_{k2}^2}\right]\frac{\sigma_{k2}^4}{(\sigma_{k2}^2-b(\mathbf{W}')\sigma_{k1}^2)^2}\right)$$
$$\int_{\mathcal{V}_1(\mathbf{W}')}x_2^2f(x_1,x_2)dx_1dx_2=2\sum_{k=1}^{2}\epsilon_k\sigma_{k2}^2\left\{1-\exp\left[-\frac{a(\mathbf{W}')}{2\sigma_{k2}^2}\right]\frac{\sigma_{k2}^2}{\sigma_{k2}^2-b(\mathbf{W}')\sigma_{k1}^2}\times\left(1+\frac{a(\mathbf{W}')}{2\sigma_{k2}^2}-\frac{b(\mathbf{W}')\sigma_{k1}^2}{\sigma_{k2}^2-b(\mathbf{W}')\sigma_{k1}^2}\right)\right\} \quad (51)$$

$$\int_{\mathcal{V}_2(\mathbf{W}')}x_1^2f(x_1,x_2)dx_1dx_2=2\left[(1-\epsilon)\sigma_{11}^2+\epsilon\sigma_{21}^2\right]-\int_{\mathcal{V}_1(\mathbf{W}')}x_1^2f(x_1,x_2)dx_1dx_2$$
$$\int_{\mathcal{V}_2(\mathbf{W}')}x_2^2f(x_1,x_2)dx_1dx_2=2\left[(1-\epsilon)\sigma_{12}^2+\epsilon\sigma_{22}^2\right]-\int_{\mathcal{V}_1(\mathbf{W}')}x_2^2f(x_1,x_2)dx_1dx_2.$$

By evaluating (50), (51) at $\mathbf{W}'=\overline{\mathbf{W}'}$ and substituting into (27), it can be proven that the stationary state is given by

$$\overline{w_{ij}^2}=2\sigma_{ij}^2=\mathrm{E}[x_j^2 \mid \mathbf{x}\in\mathcal{V}_i(\overline{\mathbf{W}'})] \qquad i,j=1,2 \quad (52)$$

provided that the following inequalities are satisfied

$$\sigma_{11}^2\gg\sigma_{21}^2, \qquad \sigma_{12}^2\ll\sigma_{22}^2$$
$$\sigma_{12}^2\sigma_{22}^2\ll\sigma_{11}^4, \qquad \sigma_{11}^2\sigma_{21}^2\ll\sigma_{22}^4, \qquad \sigma_{11}^4\ll3\sigma_{22}^4. \quad (53)$$

The first two inequalities in (53) are implied by (49). Having derived the stationary state of the L_2 LVQ neural network for the contaminated Rayleigh distribution (46), we proceed to the computation of matrix $\mathbf{B}$. For $l,n=1,2$ by using (50) and (51) we obtain

$$\frac{\partial}{\partial w_{ln}^2}\hat{F}_2(\mathbf{W}')=-\frac{\partial}{\partial w_{ln}^2}\hat{F}_1(\mathbf{W}') \quad (54)$$

and

$$\frac{\partial}{\partial w_{ln}^2}\int_{\mathcal{V}_2(\mathbf{W}')}x_m^2f(x_1,x_2)dx_1dx_2=-\frac{\partial}{\partial w_{ln}^2}\int_{\mathcal{V}_1(\mathbf{W}')}x_m^2f(x_1,x_2)dx_1dx_2 \qquad m=1,2. \quad (55)$$

Therefore, only the derivatives on the right hand side of (54) and (55) must be evaluated. The computation of the above-mentioned derivatives can be found in [39]. By evaluating the derivatives at $\mathbf{W}'=\overline{\mathbf{W}'}$ and substituting into (30) for $k,l=1,2$ and $m,n=1,2$, a symmetric matrix $\mathbf{B}$ results. Therefore, $\mathbf{B}$ is diagonalizable and it possesses real eigenvalues. Let us consider the specific example: $\sigma_{11}^2=121$, $\sigma_{12}^2=4$, $\sigma_{21}^2=9$, $\sigma_{22}^2=144$ and $\epsilon=0.2$. It can be easily verified that (53) are fulfilled. Consequently, the stationary state of the L_2 LVQ neural network is described by (52). The eigenvalues of matrix $\mathbf{B}$ can be computed by a multitude of numerical algorithms [38]. A two-step procedure has been used here. First, Householder's method is applied to reduce $\mathbf{B}$ to a symmetric tridiagonal matrix. Then QL algorithm with implicit shifts has been used to determine the eigenvalues of the real symmetric tridiagonal matrix. The eigenvalues of matrix $\mathbf{B}$ for the above-mentioned choice of parameters are listed in Table II. The time constant τ_i associated with each eigenvalue λ_i can also be found in Table II for $\alpha=0.01$.

Although in many practical cases an assumption of a Dirac-delta distribution for the initial vector of squared weight-errors would suffice (i.e., $\mathrm{E}[\mathbf{u}_i(0)]=\mathbf{u}_i(0)$), we shall compute $\mathrm{E}[\mathbf{u}_i(t)], i=1,2$ by ensemble averaging $\mathbf{U}(t)=\mathbf{Y}(t)\mathbf{U}(0)$ over an adequate number of independent random selections of the initial squared weights $w_{ij}^2(0), i,j=1,2$. We have

TABLE I
2-D VORONOI NEIGHBORHOODS

Conditions on weights			$\mathcal{V}_1(\mathbf{W}')$, $\mathcal{V}_2(\mathbf{W}')$
$w_{11}^2 < w_{21}^2$	$w_{11}^2 < w_{21}^2$	$a(\mathbf{W}') > 0$	$\{\mathbf{x} \in \mathcal{X} \mid x_1^2 < \frac{b(\mathbf{W}')}{2a(\mathbf{W}')} \wedge x_2^2 \lesssim d(\mathbf{W}')\}$
	$w_{11}^2 > w_{21}^2$	$a(\mathbf{W}') < 0$	$\{\mathbf{x} \in \mathcal{X} \mid x_1^2 > \frac{b(\mathbf{W}')}{2a(\mathbf{W}')} \wedge x_2^2 \lesssim d(\mathbf{W}')\}$
		$a(\mathbf{W}') > 0$	$\{\mathbf{x} \in \mathcal{X} \mid x_1^2 > 0 \wedge x_2^2 \lesssim d(\mathbf{W}')\}$
$w_{11}^2 > w_{21}^2$	$w_{11}^2 > w_{21}^2$	$a(\mathbf{W}') > 0$	$\{\mathbf{x} \in \mathcal{X} \mid x_1^2 < \frac{b(\mathbf{W}')}{2a(\mathbf{W}')} \wedge x_2^2 \lesssim d(\mathbf{W}')\}$
	$w_{11}^2 < w_{21}^2$	$a(\mathbf{W}') < 0$	$\{\mathbf{x} \in \mathcal{X} \mid x_1^2 > \frac{b(\mathbf{W}')}{2a(\mathbf{W}')} \wedge x_2^2 \lesssim d(\mathbf{W}')\}$
		$a(\mathbf{W}') > 0$	$\{\mathbf{x} \in \mathcal{X} \mid x_1^2 > 0 \wedge x_2^2 \lesssim d(\mathbf{W}')\}$

TABLE II
EIGENVALUES OF MATRIX B AND TIME CONSTANTS CORRESPONDING TO THEM

i	λ_i	τ_i
1	0.034589	2891.092
2	0.168021	613.417
3	0.743123	134.561
4	0.833784	119.648

TABLE III
OVERALL TIME CONSTANTS FOR THE SQUARED WEIGHTS $w_{ij}^2(t)$

Neuron i	Weight j	τ_a
1	1	125
	2	121
2	1	615
	2	1780

used 200 independent random selections of the initial squared weights in our study. The initial squared weights $w_{11}^2(0)$ and $w_{21}^2(0)$ are chosen to be uniformly distributed in an interval $\pm 2\sigma_{21}^2$ around the stationary state (52). Similarly, $w_{12}^2(0)$ and $w_{22}^2(0)$ are uniformly distributed in an interval $\pm 2\sigma_{12}^2$ around the stationary state (52). Consequently, we have (56), which is shown at the bottom of the page. The ensemble-averaged squared weight errors $\mathrm{E}[u_{ij}(t)]$ at time t over all these 200 independent trials are plotted in Fig. 1. The overall time constants τ_a for the squared weights measured using the procedure outlined above are given in Table III. By the inspection of Tables II–III, the validity of the lower and upper bounds derived in (38) can easily be deduced. It is seen the first neuron $\mathbf{w}_1' = (w_{11}^2, w_{12}^2)^T$ converges in the mean to the stationary state at a rate being close to the lower bound of (38). The convergence in the mean of the second neuron $\mathbf{w}_2' = (w_{21}^2, w_{22}^2)^T$ is slower than that of the first neuron.

Subsequently, we shall examine the convergence of the L_2 LVQ in the mean square for the contaminated Rayleigh distribution (46). To do so, the computation of matrix $\mathbf{D}$ having elements defined in (41) will be treated. The evaluation of the integrals appearing in the first three terms of (41) has already been done in (50) and (51). Thus, only the integral of the last term remains to be evaluated for $k = 1, 2$ and $m, n = 1, 2$. The evaluation of the remaining integral can be found in [39]. If the inequalities (53) are satisfied, it can be shown [39]

$$\begin{aligned}\mathrm{tr}[\mathbf{D}] &\approx 4\{(1-\epsilon)[\sigma_{11}^4 + \sigma_{12}^4] + \epsilon[\sigma_{21}^4 + \sigma_{22}^4]\} \\ &= (1-\epsilon)\mathrm{E}[\mathbf{x}'^T\mathbf{x}' \mid \mathbf{x} \in \mathcal{V}_1(\overline{\mathbf{W}'})] \\ &\quad + \epsilon\mathrm{E}[\mathbf{x}'^T\mathbf{x}' \mid \mathbf{x} \in \mathcal{V}_2(\overline{\mathbf{W}'})].\end{aligned} \tag{57}$$

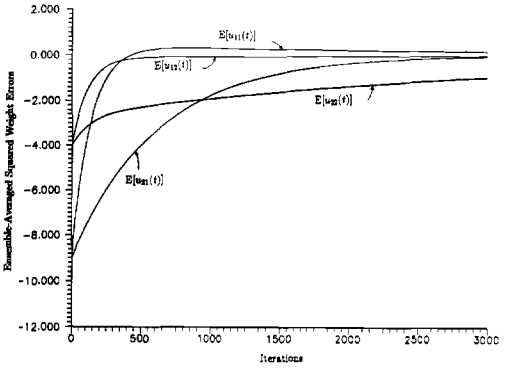


Fig. 1. Plot of ensemble-averaged squared weight errors $\mathrm{E}[u_{ij}(t)]$ $i, j = 1, 2$ versus t.

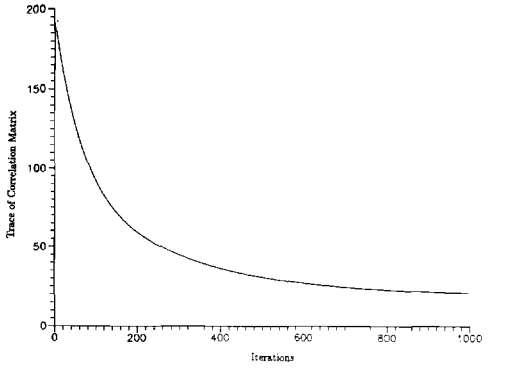


Fig. 2. Plot of the trace of the correlation matrix $J(t)$ versus t.

This trace is equal to the total power of squared input observations. In order to demonstrate the convergence of the trace of the correlation matrix, an initial correlation matrix $\mathbf{C}(0)$ has been constructed by computing the ensemble average of $\mathbf{u}_i(0)\mathbf{u}_i^T(0)$ over 200 independent random selections of the initial squared weights $w_{ij}^2(0), i, j = 1, 2$. The correlation matrix at each time instant t has been computed numerically. A plot of the trace of the correlation matrix $J(t)$ is shown in Fig. 2. As can be seen from (44), the overall time constant for $J(t)$ is expected to lie between lower and upper bounds two times smaller than those in (38). By inspecting Table II, we conclude that the lower time constant will be about 60 iterations and the upper one will be 1446 iterations for the parameter values that have been used. The measured time constant in Fig. 2 is found to be 154 iterations. Thus, inequality (44) is verified.

$$\begin{aligned} 2(\sigma_{11}^2 - \sigma_{21}^2) \le w_{11}^2(0) \le 2(\sigma_{11}^2 + \sigma_{21}^2) \qquad & 0 \le w_{12}^2(0) \le 4\sigma_{12}^2 \\ 0 \le w_{21}^2(0) \le 4\sigma_{21}^2 \qquad & 2(\sigma_{22}^2 - \sigma_{12}^2) \le w_{22}^2(0) \le 2(\sigma_{22}^2 + \sigma_{12}^2). \end{aligned} \tag{56}$$

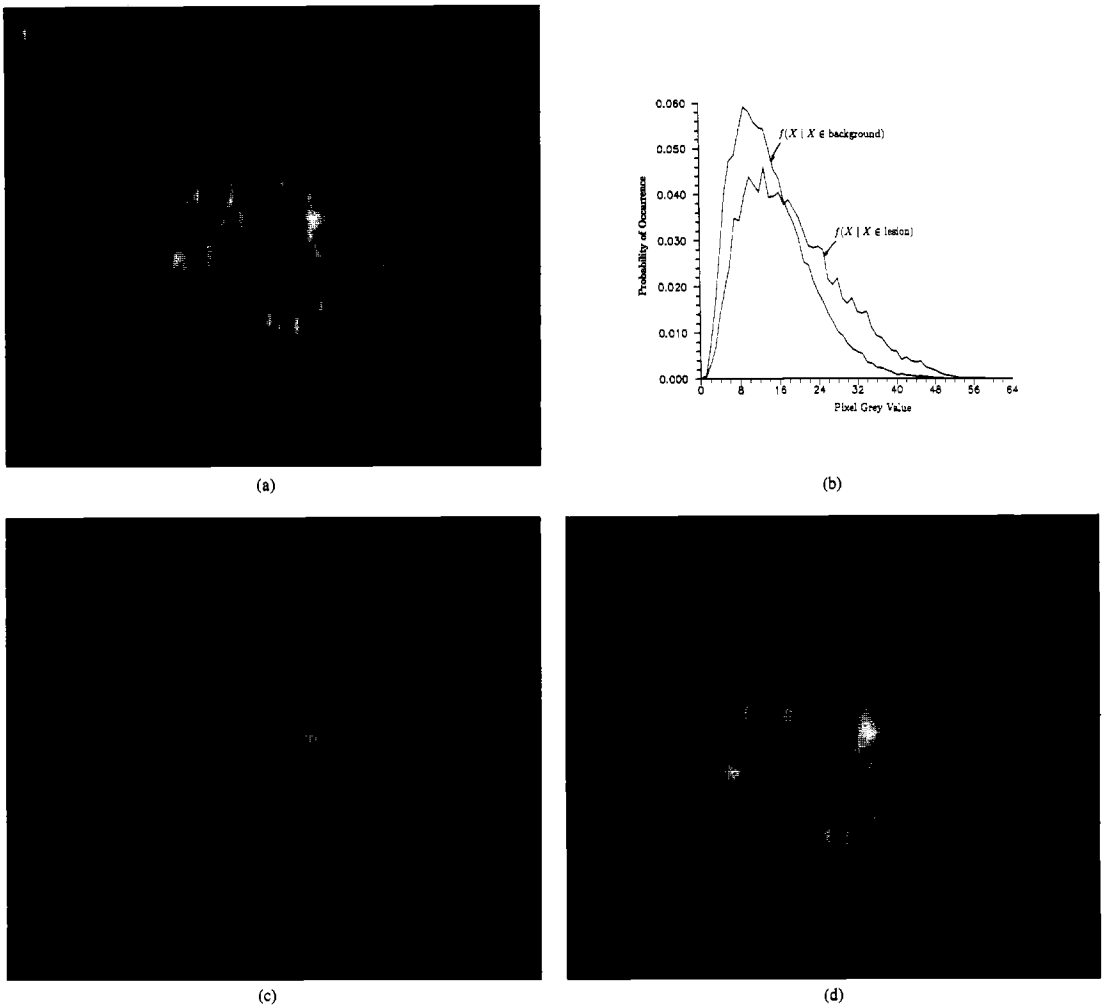


Fig. 3. Simulated US B-mode image filtering results. (a) Simulation of a homogeneous piece of tissue with a circular lesion in the middle. (b) Grey level histograms of the pixels belonging to the lesion and to the background areas. (c) Output of 5 × 5 maximum likelihood filter. (d) Output of signal-adaptive maximum likelihood filter.

IV. Experimental Results

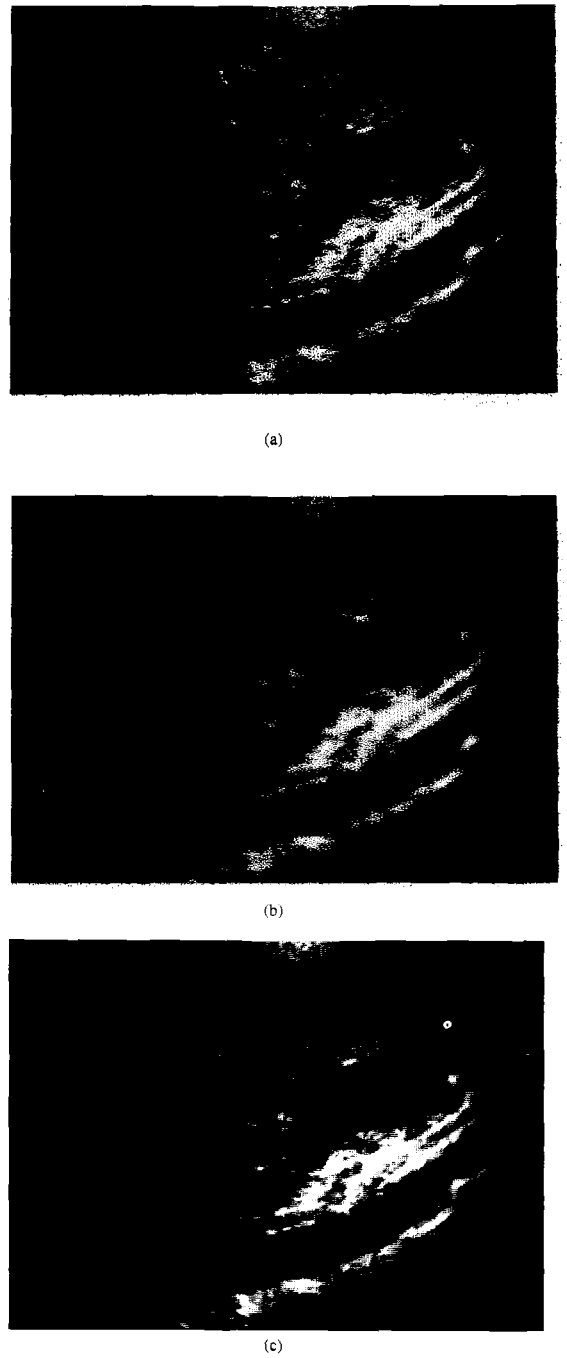
The proposed algorithms have been applied both to simulated US B-mode data and displayed US image data for speckle suppression and image segmentation. In the case of simulated B-mode data, our experiments have been performed on a set of images showing an homogeneous tissue of size 4 cm × 4 cm with a lesion in the middle of diameter 2 cm. Such a simulated US image is shown in Fig. 3(a). The lesion differs from the background only in reflection strength. The number density of scatterers in the lesion and the background is 5000/ cm^3. In addition, there is no change in second order statistics. The amplitude of reflections in the lesion is either 3 dB or 5 dB stronger than the reflection strength in the background. In the case of displayed image data, we have tested the performance of the proposed algorithms on real ultrasonic images of liver. Such an ultrasonic image is shown in Fig. 4(a).

Fig. 3(a) shows an original simulated image. The grey level histogram of the image is plotted in Fig. 3(b). It can be seen that it is very close to the Rayleigh pdf. The performance of the signal-adaptive ML filter in speckle suppression is compared to the result obtained by filtering the original image using the arithmetic mean filter, the median and the non-adaptive ML estimator of a constant original signal described in Section I.B. Due to lack of space only the filtered output of the nonadaptive ML estimator is shown in Fig. 3(c). Since signal-adaptive filters adjust their window size between 3×3 near edges and 11×11 in homogeneous regions, in order to make a fair comparison we have chosen a window size 5×5 for the

nonadaptive spatial filters. It has been found that the arithmetic mean filter blurs the image to a point where important image details such as edges are suppressed along with speckle. It has also been verified that median filter fails to comply with the requirements (e.g. noise reduction and image detail preservation) stated in the Introduction. Thus, median filter is also inappropriate for speckle suppression. ML estimator gives relatively better results from the arithmetic mean filter of the same window size. This is explained by its optimality for multiplicative Rayleigh speckle noise as analyzed in Section II.B. However, the filtered image also suffers from blurring. Image blurring may be attributed to the fact that the filter treats similarly both the lesion signal and the background. Fig. 3(d) illustrates the use of the proposed signal-adaptive ML filter for speckle suppression. Since median is not the optimal estimator for Rayleigh pdf, the signal-adaptive filter which is based on the median filter (i.e., the signal-adaptive median 14) is not expected to give better results than the proposed signal-adaptive ML filters. It is seen that the signal-adaptive ML filter performs less filtering in the middle of the image where an increase in the local signal activity (i.e., local SNR) occurs due to the presence of lesion. At the same time, it performs a considerable amount of smoothing in the background areas of the image as is desirable.

Another set of images is obtained by filtering real ultrasonic images of liver. The original ultrasonic image of liver shown in Fig. 4(a) has been filtered by a non-adaptive L_2 mean filter (Fig. 4(b)) as well as its signal-adaptive counterpart (Fig. 4(c)). Again, it is seen that the signal-adaptive L_2 mean gives the best results. Not only edges but additional diagnostically significant image details are also preserved.

A LVQ based on the L_2 mean has been created using 49 neurons at the first layer corresponding to input patterns taken from a block of 7×7 pixels. The second layer consists of 2 to 8 neurons corresponding to the output classes. A 7×7 window scans the image in a random manner to feed the network with input training patterns. During the recall phase, the 7×7 window scans the entire image in order to classify each pixel into one of p-many ($p = 2, \cdots, 8$) classes. A parametric image is created containing the class membership of each pixel. The ability of the L_2 LVQ to perform segmentation is shown in Figs. 5 and 6. Fig. 5 illustrates the classification performed by the L_2 LVQ on the simulated image. Two output classes have been used representing background and lesion respectively. Fig. 6 illustrates the segmentation of a real ultrasonic image of liver in six classes by using the L_2 LVQ. Each class is shown in Fig. 6 as a distinct image region having a grey value ranging from black to white. The more white a region is, the more important is considered to be. In general, important are the regions having strong signal activity (e.g., blood vessel boundaries, strong reflectors). Such important regions should be preserved for diagnostic purposes. Regions having rich texture are shown as light grey. In these regions, a trade off between speckle suppression and texture preservation should occur by limiting the maximal filter window size. Image regions where speckle dominates are shown dark. In these regions, speckle should be efficiently suppressed by allowing the filter window to reach its maximum size.



(a)

(b)

(c)

Fig. 4. Real ultrasonic image filtering results. (a) Real ultrasonic image of liver recorded using 3 MHz probe. (b) Output of 5×5 maximum likelihood filter. (c) Output of signal-adaptive maximum likelihood filter.

The NN approach to US image segmentation presented in this work has been compared to the following simple seg-

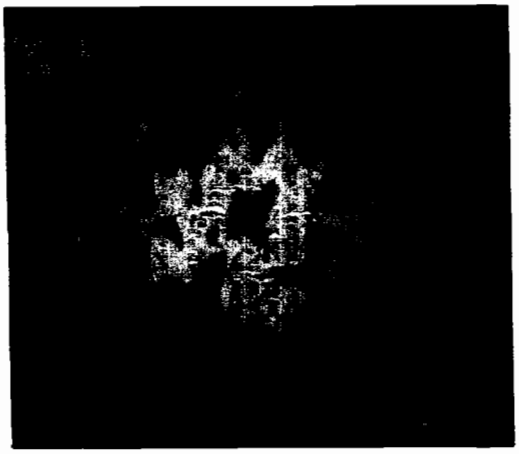


Fig. 5. Segmentation of a simulated US B-mode image by using L_2 LVQ NN

Fig. 6. Segmentation of real ultrasonic image of liver by using L_2 LVQ NN.

mentation techniques that are usually encountered in practice: (1) Image thresholding without any preprocessing. (2) Image filtering by a 7×7 median filter and thresholding the filtered image. (3) Image filtering by a 7×7 arithmetic mean filter and thresholding the filtered image. (4) Image filtering using a 7×7 ML estimator and thresholding the filtered image. We have compared the performance of the above-described strategies using the probability of detection (P_D) and the probability of false alarm (P_F) as figures of merit. The probability of detection corresponds to the percentage of pixels of the image in the lesion area that have been correctly classified. The probability of false alarm corresponds to the percentage of pixels belonging to the background of the image that were erroneously classified as belonging to the lesion. The comparison is based on the probability of detection $\hat{P}_D$ which has been calculated by linearly interpolating between the experimental values of probabilities of detection that correspond to the two closest probabilities of false alarm to the one of L_2 LVQ. The results obtained for the various classification methods are summarized in Table IV. Two simulated images have been used. These images differ in the amplitude of reflections in the lesion which is either 3 dB or 5 dB stronger than that in the background. It is seen that an almost 16.7% higher probability of detection is obtained by using the L_2 LVQ NN instead the ML estimator (4) of dimensions 7×7 in the one case. In the other, L_2 LVQ attains a 2.7% higher probability of detection than the ML estimator (4) of dimensions 7×7. In most cases, the total number of training patterns needed for sufficient training is about 10% of the image size. These pixels may not come from the same image. A set of images may be used in the learning phase resulting in a network that possesses better generalization properties. The number of iterations required in the learning phase is about 1000. This results in a computationally intensive training phase. However, training can be done off-line only once for a whole set of images, thus reducing the processing

TABLE IV
FIGURES OF MERIT FOR LESION DETECTION ON SIMULATED US B-MODE IMAGES

	3 dB				5 dB			
Method	P_F(%)	P_D(%)	Threshold	$\hat{P}_D$(%)	P_F(%)	P_D(%)	Threshold	$\hat{P}_D$(%)
Image	13.04	29.34	24	31.99	8.19	35.04	22	37.33
thresholding	15.19	32.18	23		10.12	38.13	21	
median	14.88	37.85	20	38.13	9.29	49.55	18	50.13
7×7	18.88	43.33	19		12.49	55.66	17	
arithmetic	13.78	38.95	20	41.28	7.62	52.50	18	56.15
mean 7×7	17.59	45.90	19		10.98	58.75	17	
ML estimator	13.79	40.05	19	42.39	7.81	54.30	17	57.32
7×7	17.85	47.55	18		11.65	60.80	16	
L_2 LVQ NN	15.06	59.07	-	59.07	9.60	60.05	-	60.05

time required for segmentation of the whole set. Also, as a general rule, it must be noted that the number of iterations which are necessary in the learning phase is inversely related to the number of training patterns used.

Subsequently, the performance of the modified signal-adaptive ML filter that utilizes segmentation information provided by the L_2 LVQ prior to filtering process is tested. Our purpose is to combine signal-adaptive filtering with the ability of the L_2 LVQ to segment ultrasonic images in classes representing various tissue and lesion characteristics. The result of the overall filtering process using the modified signal-adaptive ML filter is presented in Fig. 7. Parameter η has been equal to 0.5 in (19). In Fig. 7(a), it is seen that the image details are better preserved together with good speckle attenuation. For example in the upper left portion of Fig. 7(a), white blobs that are several speckle cells large are preserved in the background. Furthermore, it is seen in Fig. 7(b) that the proposed modification aids the filter in preserving better the edge information as well as acknowledging areas of the image containing valuable information that should not be filtered.

V. CONCLUSION

Two novel techniques for ultrasonic speckle suppression and ultrasonic image segmentation have been proposed. Ultrasonic

(a)

(b)

Fig. 7. Results obtained by using segmentation in conjunction with filtering. (a) Simulted US B-mode filtered image. (b) Filtered ultrasonic image of liver.

speckle suppression has been performed by using signal-adaptive filters based on the maximum likelihood estimator of the original (noiseless) signal for multiplicative noise model as well as for signal-dependent one. The ML estimator for the signal-dependent speckle model has been derived. It has been shown that it closely resembles the L_2 mean which has been proven to be the ML estimator of the original signal for the multiplicative noise model previously. Motivated by this observation, a modification of the Learning Vector Quantizer neural network called L_2 LVQ has been proposed. The convergence in the mean of the weights to the expected stationary state has been studied. The convergence of the trace of the correlation matrix of the weight-error vectors has also been considered. Lower and upper bounds for the time constants of the weights as well as for the trace of the correlation matrix have been derived. The ability of the L_2 LVQ to segment ultrasonic images in regions has been combined with signal-adaptive filtering in order to allow image detail preservation as well as maximum speckle reduction in homogeneous regions. Experimental results show that segmentation combined with filtering leads to higher speckle reduction without loss of image detail, i.e., better edge preservation and more accurate lesion detection than non-adaptive spatial filtering.

APPENDIX

Differentiating (43) with respect to t and assuming that $\mathbf{B} = \mathbf{B}^T$, we obtain

$$\dot{J}(t) = -2\alpha \mathrm{tr}[\mathbf{BC}(t)] + \alpha^2 \mathrm{tr}[\mathbf{D}]. \tag{58}$$

Let $\mathbf{\Lambda}$ be a diagonal matrix consisting of the eigenvalues of $\mathbf{B}$ and $\mathbf{Q}$ be the unitary matrix consisting of the eigenvectors associated with these eigenvalues. $\mathbf{B}$ can be expressed as $\mathbf{B} = \mathbf{Q\Lambda Q}^T$ [34]. Furthermore, let $\mathbf{K}(t) = \mathbf{Q}^T\mathbf{C}(t)\mathbf{Q}$. In general, $\mathbf{K}(t)$ is not a diagonal matrix. It can be proven that

$$\mathrm{tr}[\mathbf{BC}(t)] = \mathrm{tr}[\mathbf{Q\Lambda K}(t)\mathbf{Q}^T] = \mathrm{tr}[\mathbf{Q}^T\mathbf{Q\Lambda K}(t)] = \mathrm{tr}[\mathbf{\Lambda K}(t)] \tag{59}$$

where the following identities have been used

$$\mathrm{tr}[\mathbf{AB}] = \mathrm{tr}[\mathbf{BA}] \qquad \mathbf{Q}^T\mathbf{Q} = \mathbf{I}. \tag{60}$$

After some trivial algebraic manipulations, $\mathrm{tr}[\mathbf{\Lambda K}(t)]$ can be rewritten as

$$\mathrm{tr}[\mathbf{\Lambda K}(t)] = \sum_{l=1}^{Np}\sum_{m=1}^{Np} C_{lm}(t)\sum_{i=1}^{Np}\lambda_i Q_{li}Q_{mi}. \tag{61}$$

Consequently, $\mathrm{tr}[\mathbf{\Lambda K}(t)]$ can be bounded as follows:

$$\sum_{l=1}^{Np}\sum_{m=1}^{Np} C_{lm}(t)\lambda_{\min}\sum_{i=1}^{Np} Q_{li}Q_{mi} \le \mathrm{tr}[\mathbf{\Lambda K}(t)] \le \sum_{l=1}^{Np}\sum_{m=1}^{Np} C_{lm}(t)\lambda_{\max}\sum_{i=1}^{Np} Q_{li}Q_{mi}. \tag{62}$$

By using (60), (43) we obtain

$$\lambda_{\min} J(t) \le \mathrm{tr}[\mathbf{\Lambda K}(t)] \le \lambda_{\max} J(t). \tag{63}$$

Therefore

$$-2\alpha\lambda_{\max}J(t) + \alpha^2\mathrm{tr}[\mathbf{D}] \le \dot{J}(t) \le -2\alpha\lambda_{\min}J(t) + \alpha^2\mathrm{tr}[\mathbf{D}] \tag{64}$$

where $\lambda_{\min}, \lambda_{\max}$ denote the smallest and largest eigenvalue of matrix $\mathbf{B}$ respectively. By integrating (64), the inequalities (44) result.

ACKNOWLEDGMENT

The authors would like to thank Prof. J. M. Thijssen and Prof. J. C. Bamber for providing the simulated and clinical ultrasonic images.

REFERENCES

[1] G. W. Goodman, "Some fundamental properties of speckle," *J. Opt. Soc. Amer.*, vol. 66, pp. 1145–1150, Nov. 1976.

[2] J. G. Abbott and F. L. Thurstone, "Acoustic speckle: theory and experimental analysis," *Ultrason. Imaging*, vol. 1, pp. 303–324, 1979.

[3] S. W. Smith, R. F. Wagner, J. M. Sandrik, and H. Lopez, "Low contrast detectability and contrast/detail analysis in medical ultrasound," *IEEE Trans. Sonics Ultrason.*, vol. SU-30, no. 3, pp. 156–163, May 1983.

[4] R. F. Wagner, M. F. Insana, and D. G. Brown, "Unified approach to the detection and classification of speckle texture in diagnostic ultrasound," *Optical Eng.*, vol. 25, no. 6, pp. 738–742, June 1985.

[5] J. M. Thijssen, "Focal lesions in medical images: a detection problem, " in *Proc. NATO-ASI Mathematics Comput. Sci. Med. Imaging*, M. A. Viergever and A. Todd-Prakopek, Eds. Berlin: Springer, 1988, pp. 415–440.

[6] A. Hillion and J. M. Boucher, "A unified approach to nonlinear processing of multiplicative noise with applications to radar images, " in *Proc. V European Signal Processing Conf. (EUSIPCO '90)* (Barcelona, Spain), 1990, pp. 2027–2029.

[7] M. R. Raghuveer, S. Wear, and J. Song, "Restoration of speckle-degraded images using bispectra," in *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing (ICASSP '91)* (Toronto, Canada), 1991, pp. 3077–3079.

[8] C. Kotropoulos and I. Pitas, "Nonlinear filtering of speckle noise in ultrasonic images," in *Proc. Conf. Comput. Analysis, Image Patterns (CAIP '91)* (Dresden, Germany), Sept. 1991, pp. 118–123.

[9] C. Kotropoulos and I. Pitas, "Optimum nonlinear signal detection and estimation in the presence of ultrasonic speckle," *Ultrasonic Imaging*, vol. 14, no. 3, pp. 249–275, July 1992.

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

[11] R. C. Gonzalez and P. A. Wintz, *Digital Image Processing*, 2nd ed. Reading, MA: Addison-Wesley, 1987.

[12] D. T. Kuan, A. A. Sawchuk, T. C. Strand, and P. Chavel, "Adaptive noise smoothing filter for images with signal-dependent noise," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. PAMI-7, no. 2, pp. 165–177, Mar. 1985.

[13] J. C. Bamber and C. Daft, "Adaptive filtering for reduction of speckle in ultrasonic pulse-echo images," *Ultrasonics*, vol. 24, pp. 41–44, 1986.

[14] R. Bernstein, "Adaptive nonlinear filters for simultaneous removal of different kinds of noise in images," *IEEE Trans. Circuits Syst.*, vol. CAS-34, no. 11, pp. 1275–1291, Nov. 1987.

[15] X. Z. Sun and A. N. Venetsanopoulos, "Adaptive schemes for noise filtering and edge detection by use of local statistics," *IEEE Trans. Circuits Syst.*, vol. 35, no. 1, pp. 57–69, Jan. 1988.

[16] T. Loupas, W. N. McDicken, and P. L. Allan, "An adaptive weighted median filter for speckle suppression in medical ultrasonic images," *IEEE Trans. Circuits Syst.*, vol. 36, no. 1, pp 129–135, Jan. 1989.

[17] D. T. Kuan, A. A. Sawchuk, T. C. Strand, and P. Chavel, "Adaptive restoration of images with speckle," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-35, no. 3, pp. 373–383, Mar. 1987.

[18] A. Lopes, R. Touzi, and E. Nezry, "Adaptive speckle filters and scene heterogeneity," *IEEE Trans. Geoscience and Remote Sensing*, vol. 28, no. 6, pp. 992–1000, Nov. 1990.

[19] X. Magnisalis, C. Kotropoulos, I. Pitas, and M. G. Strintzis, "Use of fractal dimension in signal-adaptive filters for speckle reduction in ultrasound B-mode images," in *Proc. Tenth Int. Congress. Medical Informatics (MIE 91)* (Vienna, Austria), pp. 550–554, 1991.

[20] T. K. Kohonen, *Self-Organization and Associative Memory*, 3rd ed. Berlin, Heidelberg, Germany: Springer-Verlag, 1989.

[21] P. K. Simpson, *Artificial Neural Systems*. New York: Pergamon, 1990.

[22] "Special issues on neural networks, I: Theory & Modeling. II: Analysis, Techniques, & Applications," *Proc. IEEE*, vol. 78, no. 8-9, pp. 1409–1680, Sept.-Oct. 1990.

[23] R. H. Silverman and A. S. Noetzel, "Image processing and pattern recognition in ultrasonograms by backpropagation," *Neural Networks*, vol. 3, pp. 593–603, 1990.

[24] R. H. Silverman, "Segmentation of ultrasonic images with neural networks," *Int. J. Pattern Recognition*, vol. 5, no. 4, pp. 117–125, Oct. 1991.

[25] A. Loupas, "Digital image processing for noise reduction in medical ultrasonics," Ph.D. dissertation, Univ. Edinburgh, U.K., July 1988.

[26] I. Pitas and A. N. Venetsanopoulos, "Nonlinear mean filters in image processing," *IEEE Trans. Acoustics, Speech, Signal Processing*, vol. ASSP-34, no. 3, pp. 573–584, June 1986.

[27] A. Papoulis, *Probability, Random Variables and Stochastic Processes*. New York: McGraw-Hill, 1984.

[28] H. L. Van Trees, *Detection, Estimation and Modulation Theory*. New York: Wiley, 1968.

[29] T. K. Kohonen, "The self-organizing map," *Proc. IEEE*, vol. 78, no. 9, pp. 1464–1480, Sept. 1990.

[30] J. G. Proakis and D. G. Manolakis, *Introduction to Digital Signal Processing*. New York: Macmillan, 1988.

[31] H. Ritter and K. Schulten, "Convergence properties of Kohonen's topology conserving maps: Fluctuations, stability and dimension selection," *Biological Cybernetics*, vol. 60, pp. 59–71, 1989.

[32] N. G. van Kampen, *Stochastic Processes in Physics and Chemistry*. Amsterdam: North Holland, 1981.

[33] C. W. Gardiner, *Handbook of Stochastic Methods*, 2nd ed. Berlin, New York: Springer-Verlag, 1985.

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

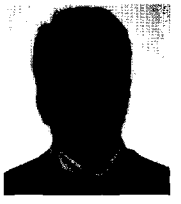
[35] F. R. Gantmacher, *The Theory of Matrices*, vols. I, II. New York: Chelsea, 1977.

[36] B. Widrow and S. D. Stearns, *Adaptive Signal Processing*. Englewood Cliffs, NJ: Prentice-Hall, 1985.

[37] S. Haykin, *Adaptive Filter Theory*. Englewood Cliffs, NJ: Prentice-Hall, 1986.

[38] J. H. Wilkinson, *The Algebraic Eigenvalue Problem*. London: Oxford, 1965.

[39] C. Kotropoulos, "Nonlinear ultrasonic image processing," Tech. Report, Univ. Thessaloniki, Greece, 1993.



C. Kotropoulos (S'88–M'92) was born in Kavala, Greece, in 1965. He received the Diploma degree with honors in electrical engineering from the Aristotle University of Thessaloniki, Greece, in 1988. He is currently working toward the Ph.D. degree in the area of nonlinear signal/image filtering.

Since 1988, he has been a research and teaching assistant in the Department of Electrical Engineering at Aristotle University. 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.

X. Magnisalis was born in Kastoria, Greece, in 1967. He received the Diploma degree in electrical engineering from the Aristotle University, Thessaloniki, Greece, in 1992. He is currently working toward the M.Sc. degree at the Ohio State University, Columbus.

During his undergraduate studies, he engaged in research at the Department of Electrical Engineering as well as at the Laboratory of Medical Informatics of the Medical School at the Aristotle University. He is currently a Graduate Research Associate at the Signal Analysis and Machine Perception Laboratory of the Electrical Engineering Department of the Ohio State University. His current research interests include image processing, computer vision, and neural networks.

I. Pitas, photograph and biography not available at time of publication.

M. G. Strintzis (S'68–M'70–SM'80), photograph and biography not available at time of publication.