A variant of Learning Vector Quantizer based on the L_2 mean for segmentation of ultrasonic images

C. Kotropoulos I. Pitas X. Magnisalis M.G. Strintzis
Department of Electrical Engineering
University of Thessaloniki
Thessaloniki 540 06, GREECE

Abstract — In this paper, the segmentation of ultrasonic images using self-organizing neural networks (NN) is investigated. A modification of Learning Vector Quantizer (called 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 variant of LVQ are studied. Experimental results show that L_2 LVQ outperforms other segmentation techniques that employ thresholding a filtered ultrasonic image with respect to the probability of detection for the same probability of false alarm in all cases.

1. Introduction

Neural networks (NN) [1, 2] 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 [2, 3, 4]. They have applied successfully in many research fields including speech and image processing, pattern recognition, data compression, control, diagnostics, knowledge representation etc. [2]. Neural networks have also found application for ultrasonic image analysis and diagnosis [5, 6]. More specifically, neural networks utilizing the backpropagation algorithm and competitive learning have been used for the localization and classification of choroidal tumors [5] as well as for enhancement and segmentation of ultrasonic images of the eye [6].

Motivated by the relative simplicity of the LVQ, its ability to work in unsupervised mode in addition to its success in image segmentation problems, a variant of the LVQ that is more suitable for ultrasonic image segmentation is proposed. More specifically, ultrasonic images suffer from a special kind of noise called *speckle*. Speckle is an interference effect caused by ultrasound (US) beam scattering from microscopic tissue inhomogeneities. Ultrasonic speckle can be modeled as multiplicative Rayleigh distributed noise or as signal-dependent Gaussian noise. The first model refers to envelope-detected US B-mode data [7]. 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.) [8]. In the case of pure multiplicative Rayleigh speckle, it has been proven that the maximum likelihood (ML) estimator of the original (noiseless) signal is the L_2 mean [9]. Furthermore, for signal-dependent Gaussian speckle, it has been shown that the ML estimator closely resembles the L_2 mean [10]. These observations 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. Such a modification will provide more accurate reference vectors for each Voronoi neighborhood and will result in a better segmentation of both ultrasonic B-mode data as well as displayed US image data. The convergence in the mean and in the mean square of the proposed L_2 LVQ NN are also studied. Therefore, the main contribution of this paper is in the derivation and study of convergence of a variant of Learning Vector Quantizer neural network based on the L_2 mean for ultrasonic image segmentation.

The outline of this paper follows. The derivation of L_2 LVQ algorithm is described in Section 2. Section 3 is devoted to the study of convergence of the proposed L_2 LVQ neural network. Experimental results are included in Section 4.

2. L_2 Learning Vector Quantizer Algorithm

Let us assume a sequence of vector-valued observations $\mathbf{x}(t) \in \mathcal{R}^N$ and a set of variable reference vectors $\{\mathbf{w}_i(t); \mathbf{w}_i \in \mathcal{R}^N, i = 1, 2, \ldots, p\}$. Learning Vector Quantizer tries to find the best-matching reference vector $\mathbf{w}_c(t)$ to $\mathbf{x}(t)$. This vector is updated. This process is repeated. After a large number of iterations, the different reference vectors tend to become specifically "tuned" to different domains of the input variable $\mathbf{x}$.

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., $\mathbf{w}'_i = (w_{i1}^2, w_{i2}^2, \ldots, w_{iN}^2)^T$. Let also $\mathbf{x}'$ denote the following vector $\mathbf{x}' = (x_1^2, x_2^2, \ldots, x_N^2)^T$. 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:

$$\varepsilon = \int_{\mathcal{X}} \| \mathbf{x}' - \mathbf{w}'_c \|^2 \; f(\mathbf{x}) \, d\mathbf{x} \tag{1}$$

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 comparing the Euclidean distances between the vector of squared

The research reported in this paper has 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.

observations and the vectors of squared weights, i.e.:

$$\| \mathbf{x}' - \mathbf{w}'_c \| = \min_i \{ \| \mathbf{x}' - \mathbf{w}'_i \| \}. \quad (2)$$

If the stochastic-gradient-descent algorithm [12] 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:

$$\begin{aligned} \mathbf{w}'_i(t+1) &= \mathbf{w}'_i(t) + \alpha(t)[\mathbf{x}'(t) - \mathbf{w}'_i(t)] \ \ \forall i \in \mathcal{N}_c(t) \\ \mathbf{w}'_i(t+1) &= \mathbf{w}'_i(t) \ \ \forall i \notin \mathcal{N}_c(t) \end{aligned} \quad (3)$$

where $\alpha(t)$ is a variable adaptation step and $\mathcal{N}_c(t)$ denotes a neighborhood around the winner. The updating equation (3) implements the unsupervised learning of the L_2 LVQ neural network. In the case of supervised learning, the vectors of squared weights are updated by using the reward-punishment procedure adopted in the so-called LVQ1 [11].

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.:

$$\mathbf{x}(t) \in \mathcal{C}_g \ \text{ if } \ \| \mathbf{x}' - \overline{\mathbf{w}'_g} \| = \min_i \{ \| \mathbf{x}' - \overline{\mathbf{w}'_i} \| \} \quad (4)$$

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

Having described the L_2 LVQ neural network, we proceed to the study of its convergence.

3. Convergence Analysis of L_2 LVQ Neural Network

A theoretical treatment of the self-organizing process is given in [13]. More specifically, the learning procedure of the LVQ has been represented by a Markov process whose states are the weight vectors $\mathbf{w}_i$. 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. The analysis presented in this paper extends the work reported in [13].

Two types of convergence, namely, the convergence in the mean and in the mean square are examined. 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. The case of a constant adaptation step $\alpha(t) = \alpha$ is considered for mathematical simplicity. Generalization for the optimal adaptation step sequence $\alpha(t) = 1/t$ [13] is also considered. We shall confine ourselves to the analysis of a single-winner L_2 LVQ network, i.e., $\mathcal{N}_c(t) = \{c\}$. Out 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 (2), i.e.,:

$$\begin{aligned} \mathcal{V}_i(\mathbf{W}') = \{ \mathbf{x} \in \mathcal{X} \subseteq \mathcal{R}^N \mid \| \mathbf{x}' - \mathbf{w}'_i \| \leq \| \mathbf{x}' - \mathbf{w}'_l \| \\ l = 1, \ldots, p, \ l \neq i \} \end{aligned} \quad (5)$$

where $\mathbf{W}' = (\mathbf{w}'^T_1 \mid \mathbf{w}'^T_2 \mid \ldots \mid \mathbf{w}'^T_p)^T$. Following the analysis in [13], the expected stationary state of the network is given by:

$$\overline{\mathbf{w}'_i} = \mathrm{E}[\mathbf{w}'_i] = \frac{\int_{\mathcal{V}_i(\overline{\mathbf{W}'})} \mathbf{x}' f(\mathbf{x}) d\mathbf{x}}{\int_{\mathcal{V}_i(\overline{\mathbf{W}'})} f(\mathbf{x}) d\mathbf{x}} \ \ i = 1, \ldots, p. \quad (6)$$

It is seen that (6) gives an implicit definition of the stationary solution of L_2 LVQ. Nonlinear equation (6) can be solved by an iterative scheme, such as the Newton method. 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 solution.

Let $\mathbf{u}_i(t)$ denote the $(N \times 1)$ vector of squared weight errors at time instant t, i.e., $\mathbf{u}_i(t) = \mathbf{w}'_i - \overline{\mathbf{w}'_i}$. The average squared weight-error vector is given by [13, 14]:

$$\mathrm{E}[\mathbf{u}_i(t)] = \mathbf{Y}(t) \mathrm{E}[\mathbf{u}_i(0)] \quad (7)$$

where the expectation is with respect to the distribution of the deviations of the squared weights from the stationary solution. $\mathbf{Y}(t)$ is the following $(Np \times Np)$ matrix:

$$\mathbf{Y}(t) = \exp(-\mathbf{B} \int_0^t \alpha(\zeta) d\zeta) \quad (8)$$

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

$$\mathbf{B} = \begin{bmatrix} \mathbf{B}_{11} & \mathbf{B}_{12} & \cdots & \mathbf{B}_{1p} \\ \vdots & & \ddots & \vdots \\ \mathbf{B}_{p1} & \mathbf{B}_{p2} & \cdots & \mathbf{B}_{pp} \end{bmatrix}. \quad (9)$$

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

$$\begin{aligned} [\mathbf{B}_{kl}(\overline{\mathbf{W}'})]_{mn} = \Bigg[w^2_{km} \frac{\partial}{\partial w^2_{ln}} \hat{F}_k(\mathbf{W}') + \hat{F}_k(\mathbf{W}') \\ \cdot \delta(k-l, m-n) - \frac{\partial}{\partial w^2_{ln}} \int_{\mathcal{V}_k(\mathbf{W}')} x^2_m f(\mathbf{x}) d\mathbf{x} \Bigg]_{\mathbf{W}' = \overline{\mathbf{W}'}} \end{aligned} \quad (10)$$

where $\hat{F}_k(\mathbf{W}') = \int_{\mathcal{V}_k(\mathbf{W}')} f(\mathbf{x}) d\mathbf{x}$ and $\delta(k-l, m-n)$ is the 2-D Kronecker-delta function, i.e.,

$$\delta(k-l, m-n) = \begin{cases} 1 & k = l \ \text{ and } \ m = n \\ 0 & \text{otherwise.} \end{cases} \quad (11)$$

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 [14]. Hopefully, the assumption that matrix $\mathbf{B}$ is symmetric is valid in many practical cases such as when the coefficient matrix $\mathbf{B}$ defined in (9)–(11) is evaluated for the contaminated Rayleigh distribution [10]. If $\mathbf{B}$ is symmetric, it is diagonalizable and possesses real eigenvalues [15]. Let λ_i, $i = 1, \ldots, q$, $q \leq Np$ be the distinct eigenvalues of matrix $\mathbf{B}$ and ρ_i be the degree of multiplicity of the eigenvalue λ_i, where $\sum_{i=1}^q \rho_i = Np$. $\mathbf{Y}(t)$ defined in (8) is a $(Np \times Np)$ matrix which can be evaluated as follows:

$$\mathbf{Y}(t) = \sum_{i=1}^{Np} y_i(t) \mathbf{B}^{i-1} \quad (12)$$

where $\mathbf{B}^0 = \mathbf{I}$ is the $(Np \times Np)$ identity matrix and $y_i(t)$, $i = 1, \ldots, Np$ are scalar functions to be determined. By applying Caley-Hamilton's theorem we obtain [16]:

$$\sum_{i=1}^{Np} y_i(t)\lambda_j^{i-1} = \exp(-\lambda_j \int_0^t \alpha(\zeta)d\zeta) \; j = 1, \ldots, q. \qquad (13)$$

It is seen that both sides of (13) are polynomials in λ. By differentiating both sides of (13) with respect to λ ν-times, where $\nu = 1, \ldots, \rho_j - 1$ and by evaluating the derivatives at each $\lambda_j \; j = 1, \ldots, q$, $(Np - q)$ additional equations result. The solution of the above-described linear set of equations determines the unknown scalar functions $y_i(t)$. In the special case that matrix $\mathbf{B}$ possesses Np distinct eigenvalues, the application of Sylvester's theorem yields the following equation for $\mathbf{Y}(t)$ [16]:

$$\mathbf{Y}(t) = \sum_{i=1}^{Np} \left(\prod_{j=1 j \neq i}^{Np} \frac{\mathbf{B} - \lambda_j \mathbf{I}}{\lambda_i - \lambda_j} \right) \exp(-\lambda_i \int_0^t \alpha(\zeta)d\zeta) \qquad (14)$$

In the following, we shall assume that the eigenvalues of matrix $\mathbf{B}$ are distinct. Therefore, we shall confine ourselves to (14). For a constant learning step $\alpha(t) = \alpha$, (14) is rewritten as follows:

$$\mathbf{Y}(t) = \sum_{i=1}^{Np} \left(\prod_{j=1 j \neq i}^{Np} \frac{\mathbf{B} - \lambda_j \mathbf{I}}{\lambda_i - \lambda_j} \right) \exp(-\lambda_i \alpha t) \qquad (15)$$

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

$$\mathbf{Y}(t) = \sum_{i=1}^{Np} \left(\prod_{j=1 \; j \neq i}^{Np} \frac{\mathbf{B} - \lambda_j \mathbf{I}}{\lambda_i - \lambda_j} \right) \frac{1}{t^{\lambda_i}} \qquad (16)$$

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

By combining (7), (15) and (16), 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., $\lambda_i > 0$. As can be seen, 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. On the contrary, it will be shown that such a bound on α can be derived if the trace of the correlation matrix of the squared weight-errors has to remain bounded.

For a constant learning step α, it is clear that $\mathbf{Y}(t)$ is a square matrix of negative exponentials. Therefore, any squared weight converges in the mean to the stationary solution (6) as a weighted sum of negative exponentials of the form $\exp(-\lambda_i \alpha t)$. The time τ_i required for each term to reach to $1/e$ of its initial value is given by:

$$\tau_i = \frac{1}{\alpha \lambda_i}. \qquad (17)$$

However, the overall time constant τ_a, defined as the time required for any average squared weight to decay to $1/e$ of its initial value, cannot be expressed in a simple closed form as (17). By using the same reasoning as in the adaptive filter literature [17], the overall time constant τ_a for any average squared weight can be bounded as follows:

$$\frac{1}{\alpha \lambda_{\max}} \leq \tau_a \leq \frac{1}{\alpha \lambda_{\min}} \qquad (18)$$

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 (9) where the $(N \times N)$ square submatrix $\mathbf{C}_{kl}(t)$, $k, l = 1, \ldots, p$ is defined by:

$$\mathbf{C}_{kl}(t) = \mathrm{E}[\mathbf{u}_k(t)\mathbf{u}_l^T(t)]. \qquad (19)$$

$\mathbf{C}(t)$ can be evaluated as follows [14]:

$$\mathbf{C}(t) = \mathbf{Y}(t) \left[\mathbf{C}(0) + \alpha^2 \int_0^t \mathbf{Y}(\zeta)^{-1} \mathbf{D} (\mathbf{Y}(\zeta)^{-1})^T d\zeta \right] \mathbf{Y}(t)^T \qquad (20)$$

where $\mathbf{C}(0)$ is the initial correlation matrix and $\mathbf{D}$ is a $(Np \times Np)$ matrix having the structure (9) 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:

$$[\mathbf{D}_{kk}(\overline{\mathbf{W}'})]_{mn} = \Bigg[w_{km}^2 w_{kn}^2 \hat{F}_k(\mathbf{W}') - w_{km}^2 \int_{\mathcal{V}_k(\mathbf{W}')} x_n^2 f(\mathbf{x}) d\mathbf{x} - w_{kn}^2 \int_{\mathcal{V}_k(\mathbf{W}')} x_m^2 f(\mathbf{x}) d\mathbf{x} + \int_{\mathcal{V}_k(\mathbf{W}')} x_m^2 x_n^2 f(\mathbf{x}) d\mathbf{x} \Bigg]_{\mathbf{W}' = \overline{\mathbf{W}'}} \qquad (21)$$

It can be seen that coefficient matrix $\mathbf{D}$ is symmetric as a result of its definition (21). Furthermore, it can be proven that matrix $\mathbf{D}$ is positive semidefinite in any case. It is known [14] that the following equation holds for the time-derivative $\dot{\mathbf{C}}(t)$ of the correlation matrix $\mathbf{C}(t)$:

$$\dot{\mathbf{C}}(t) = -\alpha[\mathbf{B}\mathbf{C}(t) + \mathbf{C}(t)\mathbf{B}^T] + \alpha^2 \mathbf{D}. \qquad (22)$$

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

$$J(t) = \mathrm{tr}[\mathbf{C}(t)] = \mathrm{E}[\sum_{i=1}^{p} \mathbf{u}_i^T(t)\mathbf{u}_i(t)]. \qquad (23)$$

It can be proven [10] that $J(t)$ can be bounded by the following inequalities:

$$\begin{aligned} &J(0)\exp(-2\alpha\lambda_{\max}t) + \alpha^2 \mathrm{tr}[\mathbf{D}]t \leq J(t) \\ &\leq J(0)\exp(-2\alpha\lambda_{\min}t) + \alpha^2 \mathrm{tr}[\mathbf{D}]t \end{aligned} \qquad (24)$$

where $J(0) = \mathrm{tr}[\mathbf{C}(0)]$. 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 (24) has practically converged to zero, i.e., at $t = 4\tau'_{\max} = \frac{2}{\alpha\lambda_{\min}}$. A sufficient condition for $J(t)$ to remain bounded is given by:

$$0 < \alpha < \frac{J_b \lambda_{\min}}{2\mathrm{tr}[\mathbf{D}]}. \qquad (25)$$

4. Experimental Results

The proposed neural network has been applied both to simulated US B-mode data as well as to displayed US image data

for image segmentation. Due to lack of space only the simulations that have been performed on a simulated image showing an homogeneous tissue of size 4 cm × 4 cm with a lesion in the middle of diameter 2 cm will be discussed. A Learning Vector Quantizer based on the L_2 mean has been created using 49 neurons at the first level corresponding to input patterns taken from a block of 7×7 pixels. The second level 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 ($p = 2, \ldots, 8$) classes. A parametric image is created containing the class membership of each pixel. Two output classes have been used representing background and lesion respectively.

The NN approach to US image segmentation presented in this work has been compared to the following simple segmentation 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 L_2 mean filter 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 probabilities of false alarm that are closest to the one of L_2 LVQ. The results obtained for the various classification methods are summarized in Table 1. It is seen that an almost 16.7% higher probability of detection is obtained by using the L_2 LVQ NN instead the L_2 mean filter of dimensions 7×7.

References

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

[2] P.K. Simpson, *Artificial Neural Systems*. Pergamon Press, 1990.

[3] Special issue on neural networks, I: theory & modeling, *Proceedings of the IEEE*, vol. 78, no. 9, pp. 1409–1544, September 1990.

[4] Special issue on neural networks, II: analysis, techniques, & applications, *Proceedings of the IEEE*, vol. 78, no. 10, pp. 1545–1680, October 1990.

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

[6] R.H. Silverman, "Segmentation of ultrasonic images with neural networks," *International journal of pattern recognition*, vol. 5, no. 4, pp. 117–125, October 1991.

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

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

[9] C. Kotropoulos and I. Pitas, "Optimum nonlinear signal detection and estimation in the presence of ultrasonic speckle," *Ultrasonic Imaging*, vol. 14, pp. 249–275, 1992.

[10] C. Kotropoulos, X. Magnisalis, I. Pitas and M.G. Strintzis, "Nonlinear ultrasonic image processing based on signal-adaptive filters and self-organizing neural networks," unpublished.

[11] T.K. Kohonen, "The self-organizing map," *Proceedings of the IEEE*, vol. 78, no. 9, pp. 1464–1480, September 1990.

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

[13] 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.

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

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

[16] F.R. Gantmacher, *The Theory of Matrices*, vol. I,II. New York: Chelsea Publishing Company, 1977.

[17] S. Haykin, *Adaptive Filter Theory*, Englewood Cliffs, N.J.: Prentice-Hall, 1986.

Table 1: Figures of Merit for Lesion Detection on a Simulated US B-Mode Image.

Method	$P_F(\%)$	$P_D(\%)$	Threshold	$\hat{P}_D(\%)$
Image	13.04	29.34	24	31.99
thresholding	15.19	32.18	23	
median	14.85	37.85	20	38.13
7 × 7	18.88	43.33	19	
arithmetic	13.78	38.95	20	41.28
mean 7 × 7	17.59	45.90	19	
L_2 mean	13.79	40.05	19	42.39
7 × 7	17.85	47.55	18	
L_2 LVQ NN	15.06	59.07	-	59.07