NONLINEAR ADAPTIVE FILTER PERFORMANCE IN TYPICAL APPLICATIONS

Peter M. Grant, S. Chen[1] and B. Mulgrew

Electrical Engineering, University of Edinburgh, Edinburgh, Scotland
[1] *now at Electrical Engineering, University of Portsmouth, Portsmouth, UK*

Abstract: This paper examines approaches to the realisation of nonlinear filters as used in signal and image processing. The design of adaptive nonlinear processors is examined and their application as adaptive equalisers to alleviate bandlimiting, distortion and interference in a typical communications channel is investigated. This paper re-examines the equalisation process as one which seeks to correctly classify the channel output into one of a finite and known alphabet of symbols encompassing the data at the channel input. The optimal solution for this classification problem is shown to be inherently nonlinear. Several nonlinear structures are examined, which allow much more complex classification boundaries, and provide greatly enhanced performance for the nonlinear filter over the more conventional linear filter. Finally the use of a nonlinear predictor is investigated for time series analysis.

Keywords: adaptive filter, adaptive equaliser, neural network, nonlinear filter, prediction methods

1. INTRODUCTION

Linear processors are bounded by the fact that the output is formed by a weighted sum of the input signal values or samples. They are a restricted class of systems but they do predominate in most electronic communication systems. The linear property means that the response to a sum of two or more inputs is equal to the sum of the responses to the individual inputs, as is implied in the principle of superposition. Systems are often not only linear but also time invariant, in that their properties such as gain etc, hopefully do not alter with time.

Many electronic components, however, only provide linear operation over a restricted range of (small) signals and when the signal amplitude increases then the system becomes nonlinear. This is clearly true in an amplifier as the signal magnitude approaches the power supply voltages then saturation in the gain stages introduces limiting or clipping and the output signal is no longer related, in a linear manner, to the input signal, Figure 1.

In electronics there are many phenomena which cause the system output to be nonlinearly related to the input. The diode detector and rectifier both have a nonlinear characteristic. Other significant nonlinear effects arise from non-Gaussian or impulsive noise. Film for recording images as well as photodetectors are nonlinear, as is the human perception system. In this latter case the logrithmic nonlinearities in human vision system are vital to achieve the large dynamic range and permit it to handle the wide range of input signals. Many other image processing operations, such as maximum entropy restoration, are nonlinear, as is the corresponding speech system which is used to represent efficiently the PCM encoded signals.
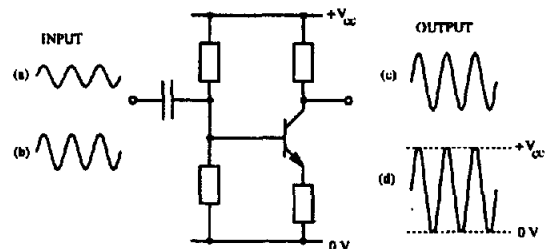

Fig. 1 Nonlinear distortion in an amplifier.

With this plethora of nonlinear operations it is not surprising that engineers have turned to using nonlinear processing techniques to more efficiently model, filter and recover these distorted signals. This paper thus discusses nonlinear processing techniques and concentrates on investigating the recent advances in nonlinear adaptive filters to process nonstationary

signal types.

2. FILTERS FOR NOISE REDUCTION

It is widely realised that if a signal is corrupted by wideband noise then the noise can be reduced by lowpass filtering. Lowpass filters can be designed into finite impulse response (FIR) filters by appropriate design and selection of the weight coefficient values, h_n, Figure 2. The relationship between the filter input, $x(k)$, and output, $y(k)$, can be summarised as:

$$y(k) = \sum_{i=0}^{N-1} h_i x(k-i) \qquad (1)$$

where N is the number of weights in the filter.

Unfortunately to be effective the passband width must be narrow and this often distorts the signal. In image processing, where a 2-dimensional filter is used, the effect of the lowpass operation removes all the high frequency components which are associated with edge features. Thus the filtered image will have blurred edges. The image processing community thus have, for some time, used nonlinear order statistic or median filters, to remove impulse noise.
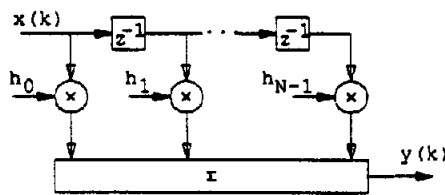

Fig. 2 Finite impulse filter.

The 1-D median FIR filter processes an odd number of pixels or samples and it rank orders the magnitudes of the samples from the largest to the smallest and outputs the middle or median value. This is clearly a nonlinear filter as the output is based on the statistics of the input data. This median filter is especially useful in impulsive noise provided that the number of noise samples encountered is less than half the number of samples or pixels being processed in the filter. Features such as edges are largely unaffected by the processing. Figure 3 shows a schematic example of a 1-D input signal as filtered in a lowpass (mean) and median filter.

More recently interest has focussed on weighted median filters (Gabbouj) and stack filters (Coyle). In the weighted median filter a non-negative integer weight is assigned to each position in the filter window, as in the FIR filter, prior to the rank order operation. These structures are particularly important when the filter is to be adaptive and the weights have

to be trained. They are used for noise reduction as shown in the processed speech waveform of Tabus, Figure 4. Other work on 2-D images (Yin) compares the mean square and mean absolute errors in lowpass, median and weighted median filters, where considerable performance improvements are obtained over the simple median filters.
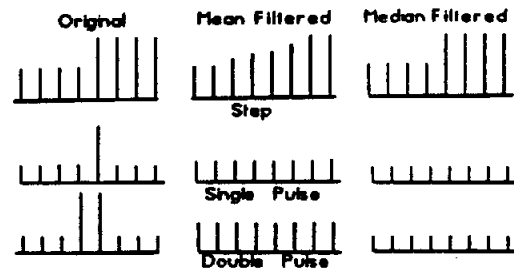

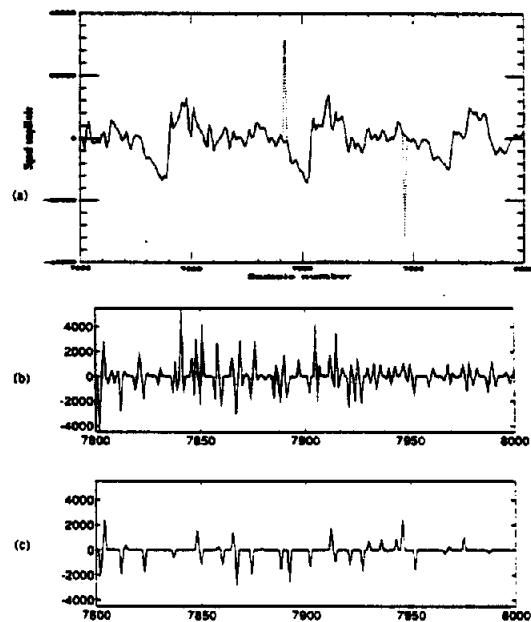**Fig. 3** Linear mean and nonlinear median filtering.

Fig. 4 Noise reduction in speech, courtesy Tabus. (a) 10 s of speech signal with added (dotted) burst errors, (b) error after 3-point median filtering as in Fig. 3, (c) reduced error after using the filter of Tabus.

Other noise reduction filters are the stack filters (Coyle). These operate by applying multiple thresholds to the input signal and binary quantising the outputs into parallel channels. These separate channels are processed in median filters and then the binary outputs are summed, at each sample value, to obtain the multi-level output. Thus the signal or image is processed in a stack of filters. This achieves a weak superposition property, not unlike a linear filter.

Stack filters with boolean operations are particularly important for efficient VLSI implementation of high speed image processing operations.

3. ADAPTIVE EQUALISATION

Adaptive channel equalisation is an old problem for which a variety of solutions have been proposed and attempted (Qureshi). There is always a commercial demand to increase the capacity of existing communications channels and to operate in more difficult environments with more severe levels of intersymbol interference (ISI), co-channel interference, additive noise and Doppler spread.

This paper re-examines the equalisation process as one which seeks to correctly classify an observed sequence (the channel output) into one of a finite and known alphabet of symbols (the data at the channel input). It demonstrates that by adopting a Bayesian approach an optimum decision boundary can be defined by certain probability density functions - effectively defining an optimal nonlinear equaliser as a mapping from observations to decisions. A geometric visualisation of classification highlights the shortcomings of the linear equaliser.

Now explore techniques for constructing the nonlinear equaliser starting with the multilayer perceptron (MLP) or neural network. Greatly enhanced performance can be achieved using this highly nonlinear structure. We demonstrate how the polynomial-based Volterra filter can also be configured as an equaliser and how its performance compares with that of the MLP. Finally, a radial basis function (RBF) structure is examined and it is show to be a more natural solution.
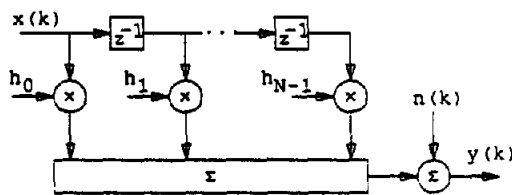

Fig. 5 Channel model.

Many digital communications channels are subject to intersymbol interference (ISI), due to restricted bandwidth and/or the presense of multipath distortion in the transmission medium. Many such channels can be characterised by a FIR digital channel filter and an additive noise source, Figure 5. The observed sequence, $\{y(k)\}$, is formed by adding Gaussian random noise to the output of the FIR channel filter, of equation (1). The relationship between the channel

input, $x(k)$, and the channel output, $y(k)$, can be summarised in the following extended equation:

$$y(k) = \sum_{i=0}^{N-1} h_i x(k-i) + n(k) \qquad (2)$$

Essentially, the equalisation problem is that of using the information present in the observed channel output vector

$$y(k) = [y(k) \; y(k-1) \cdots y(k-M+1)]^T \qquad (3)$$

to produce an estimate, $\hat{x}(k-d)$, of the channel input $x(k-d)$, as illustrated in Figure 6. The device or algorithm which performs this function is the equaliser which has order M and operates with lag (delay) d, Figure 6.
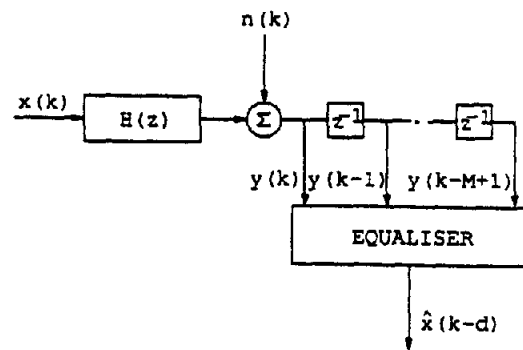

Fig. 6 Generic equaliser structure.

4. A CLASSIFICATION PROBLEM

In order to exploit the characteristics of the transmitted sequence more fully it is appropriate to highlight the finite state nature of the channel itself. To simplify the development we will assume that the transmitted sequence is binary with equal probability. For the simple case where the channel has 2 coefficients and the equaliser vector $y(k)$ has 2 elements, we can summarise all possible channel output vectors by the following equation:

$$\begin{bmatrix} y(k) \\ y(k-1) \end{bmatrix} = \begin{bmatrix} h_0 & h_1 & 0 \\ 0 & h_0 & h_1 \end{bmatrix} \begin{bmatrix} x(k) \\ x(k-1) \\ x(k-2) \end{bmatrix} + \begin{bmatrix} n(k) \\ n(k-1) \end{bmatrix} \quad (4)$$

which can be written more compactly as

$$y(k) = H x(k) + n(k) = \hat{y}(k) + n(k) \qquad (5)$$

The vector $\hat{y}(k)$ contains channel output values before noise has been added. Table 1 shows all possible channel output vectors for a simple 2 coefficient channel.

x(k)	x(k-1)	x(k-2)	ŷ(k)	ŷ(k-1)
-1	-1	-1	-1.5	-1.5
-1	-1	+1	-1.5	-0.5
-1	+1	-1	-0.5	+0.5
-1	+1	+1	-0.5	+1.5
+1	-1	-1	+0.5	-1.5
+1	-1	+1	+0.5	-0.5
+1	+1	-1	+1.5	+0.5
+1	+1	+1	+1.5	+1.5

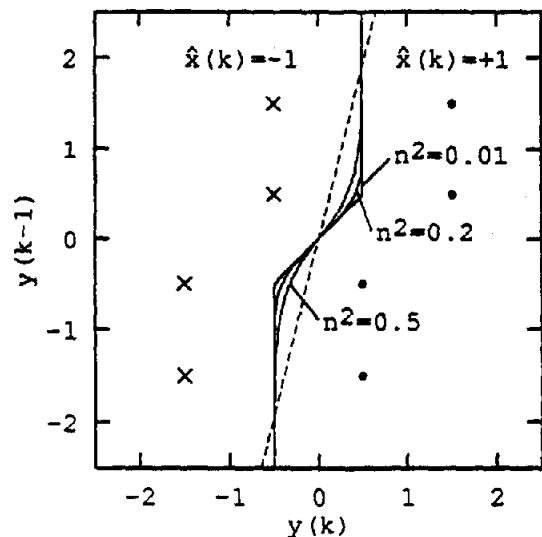Table 1 Noise-free outputs for channel $H(z) = 1 + 0.5z^{-1}$.

Fig. 7 Observation space formed by two successive outputs from channel $H(z) = 1 + 0.5z^{-1}$. Dashed line: linear decision boundary, solid curves: optimal decision boundaries with different values of noise variance n^2.

Figure 7 shows each of these 8 possible points plotted as either a × to indicate that the output vector represents an input $x(k) = -1$ or a • to represent an input $x(k) = +1$. When we add noise to the vector $\hat{y}(k)$, the 8 points become 8 clusters where the displayed points are the means or centres of the clusters. Thus the equalisation problem now involves the assigning of regions within the observation space spanned by the noisy channel output vector $y(k)$ to represent inputs of either $x(k) = +1$ or $x(k) = -1$.

The linear equaliser performs such a classification in conjunction with a decision device or slicer. The decision boundary is the locus of all values of $y(k)$ for which the output of the linear filter is zero. The decision device in this simple case is a sign function and when $y(k)$ has only two elements, it is a straight (dashed) line in Figure 7 which divides the space into two regions. All observation vectors to the right of line will be classified as indicating that $x(k) = +1$ and all points to the left of the line as $x(k) = -1$. Take for example the points $\hat{y} = [-0.5\ 0.5]^T$ and

$\hat{y} = [1.5\ 0.5]^T$. The point associated with a -1 decision is closer to the boundary than the point associated with a $+1$ decision. Therefore, in the presence of noise, there is a higher probability of a channel output vector centred on $[-0.5\ 0.5]^T$ being incorrectly detected as a $+1$ than a channel output centred on $[1.5\ 0.5]^T$ being incorrectly detected as a -1. This is clearly a non-optimum strategy.

This geometrical description of the shortcomings of the linear equaliser leads us directly to a Bayesian approach (Chen, 1990IEE). Having observed a vector $y(k)$, if the probability that it was caused by $x(k-d) = +1$ exceeds the probability that it was caused by $x(k-d) = -1$, then we should decide in favour of $+1$ and vice versa. The optimal decision boundary is the locus of all values of $y(k)$ for which the probability that $x(k-d) = +1$ given a value for $y(k)$ is equal to the probability that $x(k-d) = -1$ given the same $y(k)$. More formally, the decision boundary can be defined as the solution to a conditional probability equation, which leads to the following nonlinear equaliser function (Chen, 1990IEE) based on the difference of conditional density functions:

$$\hat{x}(k-d) = \text{sgn} (\ p_{y|x}(y(k)|x(k-d) = +1) -$$
$$p_{y|x}(y(k)|x(k-d) = -1)\ )\qquad(6)$$

The density functions and hence the equaliser can now be derived directly from the generation mechanism summarised in equation (4). The decision boundary defined by this analysis is illustrated for various levels of noise in Figure 7. For low levels of noise the boundary is piecewise linear but it becomes more smooth as the noise level increases. Clearly to equalise this channel we require to use a nonlinear filter.

The above geometric analysis can easily be applied to the general equaliser structure of Figure 6. In general, the optimal Bayesian decision boundary is a hyper-surface in the M-dimensional observation space and the realization of this nonlinear boundary requires nonlinear decision capability. The linear equaliser can however only implement a hyperplane decision boundary.

5. THE MULTILAYER PERCEPTRON

In practice we do not know the channel impulse response and hence cannot directly form the function defined by equation (6). To construct this function adaptively using training data, we might first apply a very general nonlinear structure known as the multilayer perceptron (MLP) (Lippmann) to the problem.

The basic building block of the MLP is the single neuron or node which is shown in Figure 8. A node receives a number of inputs $x_1, \cdots, x_n$, say, which are then multiplied by a set of weights $w_1, \cdots, w_n$ and the resultant values are summed. A constant θ, known as the node threshold, is added to this weighted sum of inputs, and the output of the node is obtained by evaluating a nonlinear function, f, of the total. The output mapping function is the operation which makes this processor nonlinear. We focus our attention on nodes where the node activation function, f, is defined by

$$f(x) = (1 - e^{-x})/(1 + e^{-x}) \qquad (7)$$
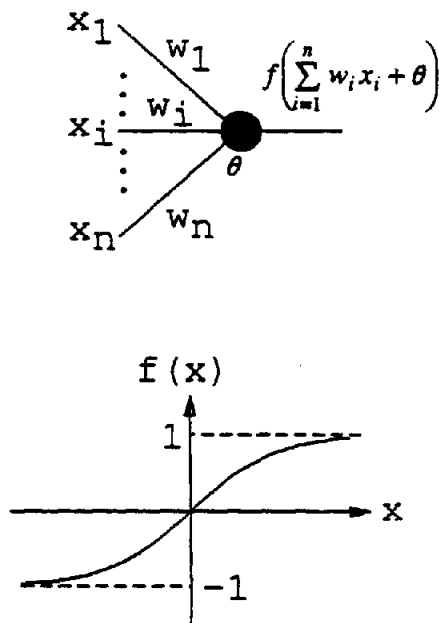
the graph of which is also shown in Figure 8.

Fig. 8 Node structure and activation function.

In the MLP a number of nodes are arranged in layers, as depicted in Figure 9. A multi-dimensional input is passed to each node in the first layer. The outputs of the first layer nodes then become inputs to the nodes in the second layer, and so on. The output of the network is therefore the outputs of the nodes lying in the final layer. Thus, weighted connections exist from a node to every node in the succeeding layer, but no connections exist between nodes in the same layer. Since an equaliser usually has one output the network has a single node in the output layer.

MLPs which have three layers are essentially capable of forming any desired decision region (Gibson, 1990), and it is this property which makes them attractive as nonlinear equalisers. When required to

learn a task adaptively, MLPs can be trained by means of the back-propagation (BP) algorithm (Rumelhart). This is a simple stochastic gradient-descent algorithm similar to that used in linear equalisers (Qureshi).

The major difficulty with the MLP is that training is essentially a nonlinear optimisation problem. The mean squared error surface is multi-modal and hence it is extremely difficult to design gradient type algorithms which are guaranteed to find the global minimum under all input signal conditions, hence the next algorithm development.
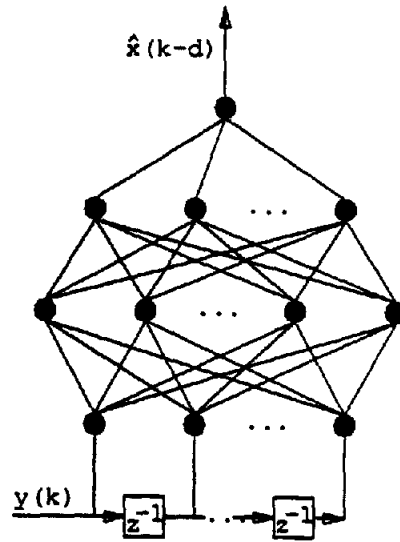

Fig. 9 Multilayer perceptron equaliser.

The slab algorithm (Hassell Sweatman *et al.* 1994) is a deterministic design method for constructing a 2 layer MLP from McCulloch and Pitts units i.e. the node activation function of equation (7) is $f(x) = \mathrm{sgn}(x)$ and is such that the node output is either $+1$ or -1. The output of the ith node in the hidden layer is typically:

$$\mathrm{sgn}(\,\mathbf{y}^T(k)\,\mathbf{w}_i - \theta_i\,)$$

where $\mathbf{w}_i$ is the vector of weights on the input to the ith node and θ_i is the threshold. The algorithm requires knowledge of the noise free states or equivalently the channel impulse response and application of linear programming techniques.

Figure 10 illustrates the application of the algorithm to a simple channel equalisation problem with a complicated decision space. The noise free states associated with transmitting a $+1$ or -1 are shown on Figure 10(a). The hyperplane A is chosen such that all states associated with a -1 are to the right of it. Its companion, B, is chosen such that all points to its left

are associated with a +1. If A and B are the same hyperplane then the states are linearly seperable, as the hyperplanes are identical and the common hyperplane defines the coefficients of a linear equaliser. If the states are not linearly seperable the space between the two lines is "slab 1". The width of the slab is chosen to minimise the number of states within it. The hyperplanes A and B define two nodes on the MLP structure illustrated on Figure 10(b). The weights on the input to the second layer of the network are now taken from a geometric series. In this case any decision made by nodes A and B will dominate the final decision and they have significant weight values ($\frac{1}{2}$). Thus if the input vector $y(k)$ is to the left of hyperplane A an output decision of +1 will result. If the vector is to the right of B an output decision of −1 will result. Finally if the vector is within slab 1, nodes A and B will contribute a total of 0 to the output node and will not affect the final decision.

The next step is to consider only states within slab 1 and repeat the process. All points associated with a +1 are above hyperplane D, all points associated with a −1 are below hyperplane C and the number of states within slab 2 has been minimised. Hyperlanes C and D give us nodes C and D in the network. Finally within slab 2 the states are linearly separable with the further hyperplane E giving node E in the network with the smallest weight values on the output node.

In comparing the slab algorithm, against the more conventional MLP, the slab algorithm is more predictable in training to achieve the required weight solution but it may suffer a minor performance degradation.

(a)

KEY: • $x(k-l) = +l$

          o $x(k-l) = -l$
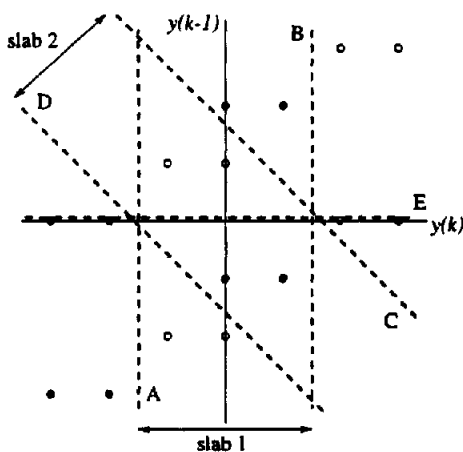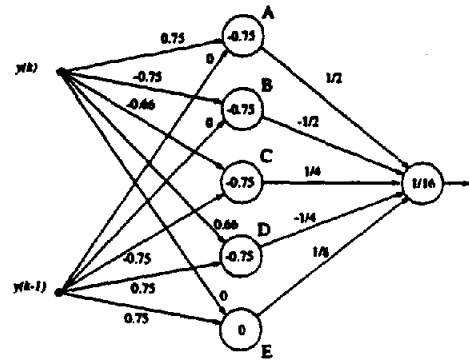
(b)

Fig. 10 MLP constructed by the Slab algorithm for channel $H(z) = 0.333 + 0.667z^{-1} + z^{-2}$: (a) state structure and resultant hyperplanes; (b) MLP architecture.

6. THE VOLTERRA SERIES

The classical approach to dealing with nonlinear systems is the Volterra series (VS) expansion. Essentially, it is a modification to a Taylor series to include a time dispersive or memory element. Using a VS we can construct a nonlinear equaliser of the form (Chen, 1990IEE):

$$f_v(y(k)) = \sum_{i=0}^{M-1} w_i y(k-i) + \sum_{i=0}^{M-1}\sum_{j=i}^{M-1} w_{ij} y(k-i)y(k-j)$$

$$+ \sum_{i=0}^{M-1}\sum_{j=i}^{M-1}\sum_{l=j}^{M-1} w_{ijl} y(k-i)y(k-j)y(k-l) + + + \quad (8)$$

$$\hat{x}(k-d) = \text{sgn}(f_v(y(k))) \quad (9)$$

The coefficients w_i, w_{ij}, w_{ijl}, etc are known as the Volterra kernals.

In theory the number of terms in equation (8) can be infinite. However in practice only a finite number of terms can be implemented. Consider a simple VS equaliser where $M = 2$ and the polynomial expansion is truncated to degree-2 terms:

$$\hat{x}(k-d) = \text{sgn}(w_0 y(k) + w_1 y(k-1) +$$

$$w_{00} y^2(k) + w_{01} y(k)y(k-1) + w_{11} y^2(k-1)) \quad (10)$$

The structure of the equaliser defined by equation (10) is illustrated in Figure 11. The advantage inherent in the VS equaliser over the MPL is that it is linear in the parameters. The equaliser can be trained by using standard linear adaptive algorithm to update the kernal vector w_v. A drawback of the VS equaliser is that the number of kernals increases exponentially as the degree or the order M increases.
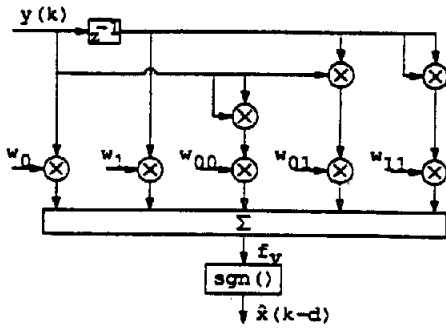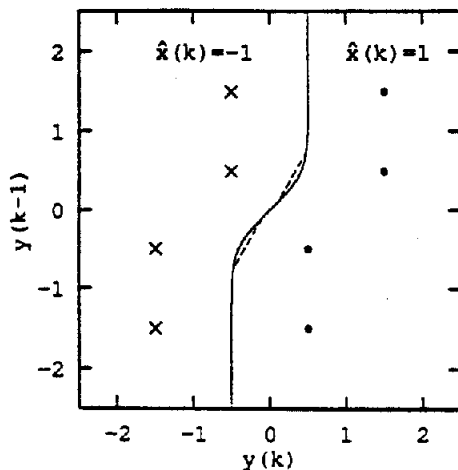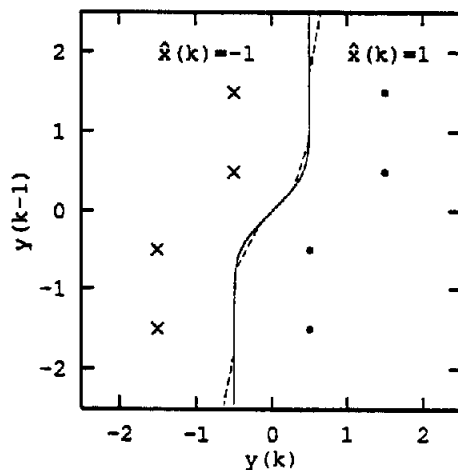
Fig. 11 Volterra equaliser of order 2 and degree 2.

(a) Dashed curve: MLP equaliser.

(b) Dashed curve: VS equaliser.

Fig. 12 Comparison of decision boundaries. Channel $H(z) = 1 + 0.5z^{-1}$, solid curve: optimal decision boundary.

Figure 12 illustrates the performance of the straightforward MLP and VS equalisers when used to equalise a simple channel with transfer function $H(z) = 1 + 0.5z^{-1}$. The power of the additive noise,

n^2 in Figure 5 and equation (2), was 0.2, representing a signal to noise ratio (SNR) of approximately 8 dB. The MLP has a $2-9-5-1$ configuration. That is, the observation vector $y(k)$ has 2 elements corresponding to the two tap channel transfer function, the first hidden layer has 9 nodes, the second hidden layer has 5 nodes and the output layer has 1 node. The VS equaliser has a structure of order 2 and degree 3. The decision boundaries of these two non-linear equalisers, obtained after a training sequence of 300 samples, are compared with the optimal decision boundary in Figure 12(a) and (b) respectively. Performance of the MLP and VS equalisers for more realistic channels and equaliser configurations have extensively been studied (Chen, 1990IEE; Gibson, 1991SP).

7. THE RADIAL BASIS FUNCTION NETWORK

In this section a third adaptive nonlinear structure, namely the radial basis function (RBF) network of Broomhead, is described. The RBF network is ideal for equalisation application because it has an equivalent structure to the optimal Bayesian equalisation solution (Chen, 1993NN). The schematic of the RBF network is depicted in Figure 13. The overall response of the network can be summarised in the following equation:

$$f_r(\mathbf{y}(k)) = \sum_{i=1}^{n_r} w_i \phi_i \qquad (11)$$

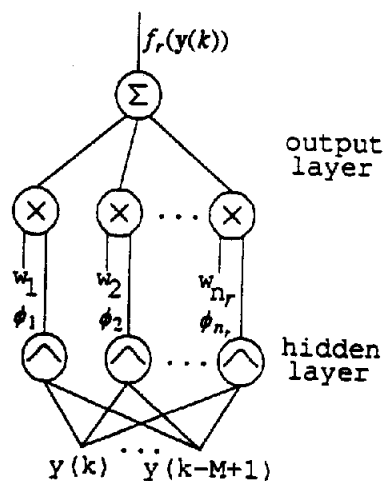$$= \sum_{i=1}^{n_r} w_i \phi(\|\mathbf{y}(k) - \mathbf{c}_i\|^2 / \rho_i)$$

Fig. 13 Radial basis function network.

c_i are the RBF centres which have the same dimensional as the input vector $y(k)$, $\|\cdot\|$ denotes the Euclidean norm, ρ_i are the positive constants known as the widths, and $\phi(\cdot)$ is the basis function. To use the RBF network for detecting the channel input $x(k-d)$, the output of the network, f_r, is passed through a decision slicer.

The relationship between the RBF network and the Bayesian equalisation solution, equation (6), can be given explicitly (Mulgrew, 1996). The number of centres, n_r, is equal to that of the noise-free channel output vector $\hat{y}$, and c_i are in fact these vectors. The weights w_i are all known and they are the corresponding scaling factors of the conditional density functions in equation (6). The widths ρ_i are controlled by the noise variance n^2, and are usually set at $\rho_i = 2\sigma_n^2$, while $\phi(\cdot)$ is the noise probability density function which is usually Gaussian. When these conditions are met, the RBF network realizes precisely the Bayesian equalisation solution.

To realize the optimal Bayesian solution using the RBF network, one thus needs to identify the centres or the noise-free channel output vectors. Chen (1993) conveniently achieved this using two alternative schemes. The first method identifies the channel model using standard linear adaptive algorithm and then calculates the noise-free vectors from the generation mechanism, $H\,x(k)$, of equation (5). The second method estimates these vectors or centres directly using a supervised and unsupervised clustering algorithms (Chen, 1993NN and 1993SP).
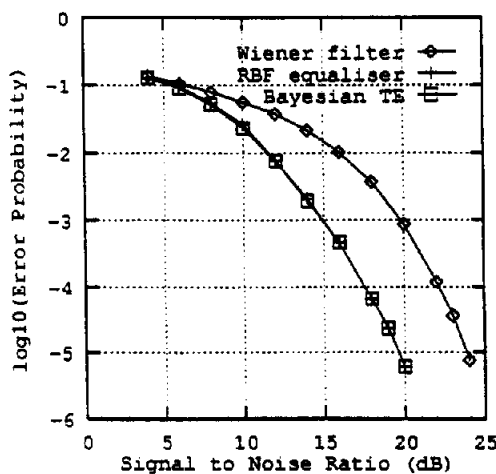

Fig. 14 Performance on the fixed channel $H(z) = 0.3482 + 0.8704z^{-1} + 0.3482z^{-2}$.

For the simple channel and the equaliser configuration given in Figure 13, 60 training samples were used to estimate the RBF centres based on the clustering scheme. The resulting RBF equaliser produces a decision boundary which is indistinguishable from the optimal Bayesian decision boundary. The bit error rate of the adaptive RBF equaliser was further compared with that of the optimal Bayesian equaliser for the channel with impulse response $H(z) = 0.3482 +0.8704z^{-1} + 0.3482z^{-2}$ under a variety of SNR values. The equaliser had a structure of $M = 4$ and $d = 1$. The adaptive RBF equaliser used the least mean square algorithm to identify the channel model with a training sequence of 90 samples. The bit error rate curves of the adaptive RBF and Bayesian equalisers are depicted in Figure 14. The third error rate curve in Figure 14 was obtained by the Wiener filter which is the theoretical performance bound of the linear equaliser. From Figure 14, it is seen that the adaptive RBF equaliser realizes the optimal performance and is far superior to the linear equaliser.

Callender(1994) has also compared the RBF, MLP and VS equalisers and added distance classification techniques based on the Mahalanobis distance metric. The modified Mahalanobis (Callender, 1992) adds noise to the covariance update recursion to avoid convergence problems.

The intimate link between the RBF network and the Bayesian equaliser makes the RBF design an attractive solution to equalisation problems. The performance of the RBF equaliser is superior to the MLP and VS equalisers and it needs a much shorter training period than these other two nonlinear equalisers. Moreover the two adaptive schemes for the RBF equaliser are guaranteed to converge to the optimal Bayesian solution.

8. THE DECISION FEEDBACK EQUALISER

A powerful technique to improve equalisation performance is to use decision feedback (Qureshi). A generic decision feedback equaliser (DFE) is illustrated in Figure 15, where M and n_b are known as the feedforward and feedback orders respectively. This structure is obtained by expanding the equaliser input vector $y(k)$ of Figure 5 to include the past detected symbols $\hat{x}_b(k-d)=[\hat{x}(k-d-1)\cdots\hat{x}(k-d-n_b)]^T$. The conventional DFE (Qureshi) is based on a linear filtering of the expanded equaliser input vector $[y^T(k)\ \hat{x}_b^T(k-d)]^T$.

Similar to the discussion in section 4, we can view the equalisation process defined in Figure 15 as a classification problem and derive its optimal solution (Chen, 1993SP & 1992). The three nonlinear struc-

tures given previously can then be applied to adaptively implement this Bayesian solution. In particular, the RBF design has many advantages. It provides a clear explanation of how decision feedback improves performance, as well as reducing the computational complexity.
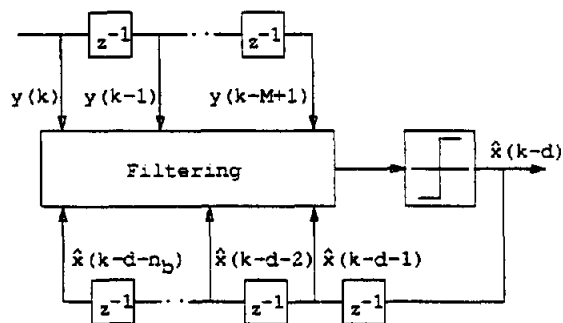

Fig. 15 A generic decision feedback equaliser.

Figure 16 shows results on a 4 symbol quadrature amplitude modulated (4-QAM) system, comparing a conventional DFE detector, a Bayesian (i.e. RBF) DFE and a maximum likelihood Viterbi algorithm (MLVA) equaliser, with delays of 2 & 10 symbols. The Bayesian is superior to the conventional detector by 3 dB but the MLVA is best with longer delay of 10 on this stationary channel, as it is able to perform a more accurate channel estimate.
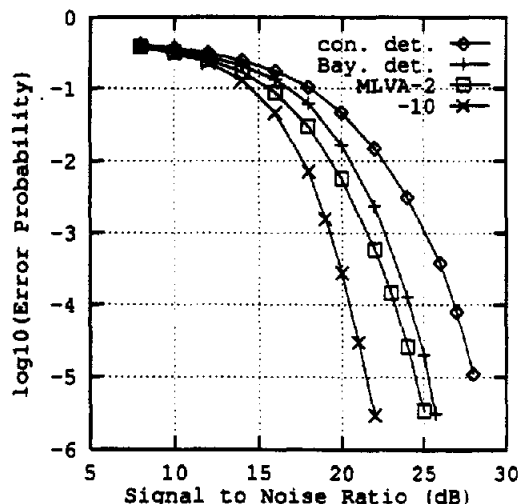
design using a simulated mobile radio nonstationary (fading) channel. The channel input $x(k)$ is again 4-QAM and the symbol rate is 300 kHz. The transmission pulse has a raised-cosine characteristic with a rolloff factor of 0.5 and is split equally between the transmitter and receiver filters. A multipath Rayleigh fading channel is simulated with a Doppler frequency of 100 Hz, and the channel dispersion spans 4 symbol durations. The time-varying channel thus has a transfer function

$$H(z) = h_0(k) + h_1(k)z^{-1} + \ldots\ldots + h_4(k)z^{-4} \quad (12)$$

The time-varying taps $h_i(k)$ and the channel output $y(k)$ are complex-valued.

Three adaptive equalisers, namely the conventional DFE, the maximum likelihood sequence estimator (MLSE) (Forney) and the Bayesian DFE based on the RBF design, are investigated. The conventional DFE is fractional spaced (FS), that is, the channel output is sampled at a rate faster than the symbol rate. It has nine half-symbol spaced feedforward terms and 4 feedback terms. The MLSE is implemented as a Viterbi detector with a fixed decision delay of 6 symbol periods and equipped with a least mean square channel estimator. The Bayesian DFE based on the RBF design has a structure of $d = 2$, $M = 3$ and $n_b = 4$, and it employs a least mean square channel estimator to adaptively update the RBF centres. The symbol error rates of these three adaptive equalisers are depicted in Figure 17.
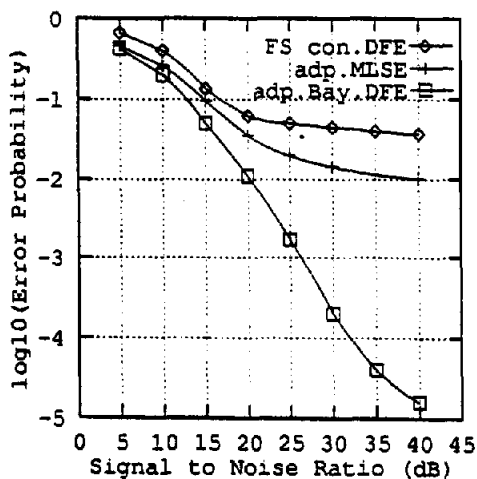

Fig. 16 DFE performance comparisons on the same channel as used in Fig. 14.

Chen (1993SP) has also demonstrated adaptive performance of the Bayesian DFE based on the RBF

Fig. 17 Performance comparison for a multipath fading mobile radio channel with 100 Hz Doppler.

From Figure 17, it can be seen that the adaptive Bayesian DFE based on the RBF design is far superior over the adaptive MLSE. This is very interesting since the theoretical performance of the Bayesian DFE is inferior to that of the MLSE. The equalisation

configurations of Figures 6 and 16 are known as the symbol-by-symbol equaliser where detection of the channel input sequence is on a symbol-by-symbol basis. The MLSE on the other hand is the optimal solution for a very different class of sequence-estimation equalisers. It provides the lowest error rate attainable for any equaliser when the channel is known. However, for fast time-varying channels such as moble radio fading channels, the adaptive MLSE suffers serious performance degradation due to channel tracking errors. In contrast the adaptive Bayesian DFE based on the RBF design is very robust and suffers a much smaller degradation in the time-varying environment. A more extensive simulation study is presented in (Chen, 1995).

It has also been convincingly demonstrated how the RBF can be deployed to realise new designs of blind equalisers with equally significant performance results (Chen, 1994ICC).

9. SIGNAL PREDICTION

These nonlinear processing techniques are equally applicable to signal prediction (Mulgrew, 1992) and system modelling (Chen, 1992IJC) as well as equalisation. Application of the RBF model to nonlinear signal prediction has been regularly demonstrated by analyzing for example the time series of the annual sunspot numbers. The annual sunspot numbers for the years 1700-1979 are plotted as 280 observations in Figure 18.
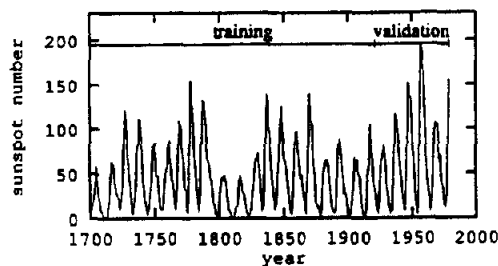

Fig. 18 Sun spot time series data.

The sunspot data from 1700 to 1920 were used for training and the data from 1921 to 1979 were employed for evaluating predictive accuracy. Because this time series is nonstationery, a shorter validation set, 1921-1955, is also used for evaluation of the prediction accuracy. Many researchers have fitted various predictors or models to the sunspot time series. These include the linear autoregressive model and the bilinear model (Gabr), the threshold autoregressive model (Tong), the polynomial or Volterra-series (Chen, 1989), and the one-hidden-layer sig-

moid network model (Chen, 1990IJC; Weigend). Chen (1994JASE) has also fitted an RBF model to this sunspot time series.

Because this is a scalar time series, Chens (1994JASE) RBF network employed a single output node. The nonlinearity of hidden nodes was chosen to be the thin-plate-spline function. The desired value for the predictor lag was derived by fitting RBF predictors of different lags and choosing one that gives the best result. For the sunspot time series, it is found that a lag of 8 was an appropriate choice. The one-step RBF predictor

$$\hat{y}(t + 1 \mid t) = f_r (y(t), y(t - 1), \cdots y(t - 7)) \quad (13)$$

was fitted to the training data set. The evaluation of the fitted RBF predictor is then done by examining the prediction error and the normalized prediction variance was further used to evaluate the fitted RBF model.
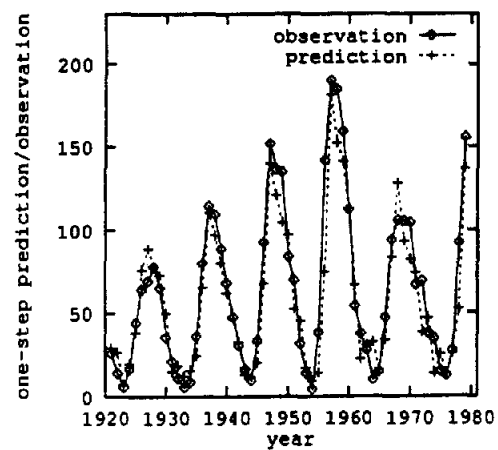

Fig. 19 RBF predictor fitted to Fig. 18 validation data with the OLS training algorithm.

The training set has 221 samples and therefore there are 221 candidates for the RBF centres. The orthogonal least squares (OLS) algorithm was initially employed to construct a one-step RBF predictor. As expected, when the OLS algorithm added more centres to the selected RBF network, the accuracy of the model over the training set continued to improve. After 25 centres had been selected, the normalised variance over the validation set began to increase as more centres were added in to the network. This suggested that the selection procedure should be terminated at the 25th step, giving rise to a RBF predictor of 25 centres. Figure 19 shows the one-step predictions over the validation set, superimposed on the time series observations.

Clustering and least squares (LS) algorithm was also employed to fit a one-step RBF predictor to the

sunspot time series (Chen 1994JASE). The number of hidden nodes was again chosen to be 25. The training input data were passed through the clustering algorithm five times with the corresponding adaptive gain being 0.2, 0.1, 0.05, 0.025 and 0.01. After the centres had been obtained, the weights were then learnt using the recusive least squares (RLS) algorithm. Repeated training was performed by passing the data through the RLS algorithm five times. The one-step predictions over the years 1921-1979 are again superimposed on the original time series observations in Figure 20. Comparing Figure 19 with Figure 18, it can be seen that the predictor obtained by the OLS algorithm has better short-term predictions than the predictor fitted by the clustering and RLS algorithm, but the long-term predictions of the latter are better than those of the former!
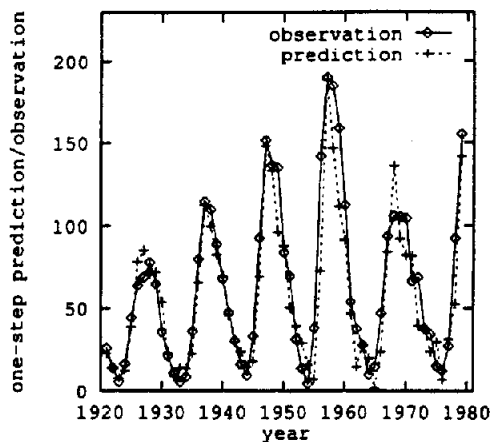

Fig. 20 RBF predictor fitted to Fig. 18 validation data with clustering and the RLS algorithm.

Some time ago Chen (1989) fitted a subset polynomial model to the sunspot time series and showed that it had better performance than the linear and bilinear models of Gabr. For the sake of comparison, Chen selected the same predictor lag of 9 as used in the linear and bilinear models. Weigend has also constructed a sigmoid network model for the sunspot time series and demonstrated its superior performance over the threshold model of Tong. They chose a predictor lag of 12 because this was the value used in the threshold model. The overall multi-step prediction performance of the RBF model obtained in Chens (1994JASE) study was better than either those of the subset polynomial model or the sigmoid network model. It is worth pointing out that only a modest number of centres were used in the Chen RBF model. Weigend surprisingly commented that the RBF approach was inefficient and required a very large number of centres. Chens more recent results clearly suggest otherwise! Chen (1994JASE) also describes other system modelling simulations using the RBF network.

10. CONCLUSIONS

This paper has attempted to briefly review the application of nonlinear filter structures to noise reduction, channel equalisation and prediction problems. These structures show considerable promise for improved noise reduction and the nonlinear adaptive filters give significantly improved equaliser performance over linear filter approaches. Recent equaliser research has focused on the radial basis function design incorporating decision feedback, and this provides a practical and powerful equalisation scheme which even outperforms the maximum likelihood sequence-detection in a highly nonstationary environment.

The results presented here are typical of those received from nonlinear filter designs, and hence the RBF is now assuming a prominant position as one of the most attractive nonlinear filter design approaches. As a result of these excellent results the RBF is now being applied to other communications processor functions such as interference cancellation (Mulgrew, 1995) and bearing estimation.

11. ACKNOWLEDGEMENTS

12. REFERENCES

D.S. Broomhead and D. Lowe, "Multivariable functional interpolation and adaptive networks," *Complex Systems*, Vol.2, pp.321-355, 1988.

C.P. Callender and C.F.N. Cowan, "Use of cluster classification for channel equalisation at high SNR", *IMA Mathematics in Signal Processing* Conference Proceedings, Warwick, December 1992.

C.P. Callender and C.F.N. Cowan, "A comparison of 6 different nonlinear equaliser techniques for digital communications systems", *Proceedings VII European Signal Processing Conference*, EUSIPCO-94, pp. 1524-1527, Edinburgh, Sept 1994.

E.J. Coyle and N.C. Gallacher, "Stack filters and neu-

ral networks", *IEEE Int. Symposium Circuits and Systems*, pp. 995-998, May 1989.

S. Chen and S.A. Billings, "Modelling and analysis of non-linear time series", *Int. Journal Control*, Vol. 50, No. 6, pp. 2151-2171, 1989.

S. Chen, S.A. Billings and P.M. Grant, "Non-linear systems identification using neural networks", *Int. Journal Control*, Vol. 51, No. 6, pp. 1191-1214, 1990.

S. Chen, G.J. Gibson and C.F.N. Cowan, "Adaptive channel equalisation using a polynomial-perceptron structure," *IEE Proceedings, Part.I*, Vol.137, No.5, pp.257-264, 1990.

S. Chen and S.A. Billings, "Neural networks for non-linear dynamic system modelling and identification," *Int. Journal Control*, Special Issue on Intelligent Control, Vol.56, No.2, pp.319-346, 1992.

S. Chen, B. Mulgrew and P.M. Grant, "A clustering technique for digital communications channel equalisation using radial basis function networks," *IEEE Trans. Neural Networks*, Vol. 4, No. 4, pp. 570-579, July 1993.

S. Chen, B. Mulgrew and S. McLaughlin, "Adaptive Bayesian equaliser with decision feedback," *IEEE Trans. Signal Processing*, Vol. 41, No. 9, pp. 2918-2926, 1993.

S. Chen, "Radial basis functions for signal prediction and system modelling", *Journal of Applied Science & Engineering*, Vol. 1, No. 1, June 1994.

S. Chen, S. McLaughlin, B. Mulgrew and P.M. Grant, "Adaptive Bayesian decision feedback equaliser incorperating cochannel interference suppression", *in proceedings IEEE Int. Conference Communications*, Vol. 1, pp. 530-533, May 1994.

S. Chen, S. McLaughlin, B. Mulgrew and P.M. Grant, "Adaptive Bayesian decision feedback equaliser for dispersive mobile radio channels," *IEEE Trans. Communications*, Vol.43, No.5, pp.1937-1946, May 1995.

G.D. Forney, "Maximum-likelihood sequence estimation of digital sequences in the presence of intersymbol interference," *IEEE Trans. Information Theory*, Vol.IT-18, No.3, pp.363-378, 1972.

M. Gabbouj, E.J. Coyle and N.C. Gallacher "An Overview of median and stack filters", *Circuits, Systems Signal Processing*, Vol. 11, No. 1, pp. 7-45, 1992.

M.M. Gabr and T. Subba Rao, "The estimation and prediction of subset bilinear time series models with applications", *Journal Time Series Analysis*, Vol. 2, pp. 155-171,

G.J. Gibson and C.F.N. Cowan, "On the decision

region of multi-layer perceptrons," *Proc. IEEE*, Vol.78, No.10, pp.1590-1594, 1990.

G.J. Gibson, S. Siu and C.F.N. Cowan, "The application of nonlinear structures to the reconstruction of binary signals," *IEEE Trans. Signal Processing*, Vol.39, No.8, pp.1877-1884, 1991.

C.Z.W. Hassell Sweatman, G.J. Gibson and B. Mulgrew, "A Constructive Algorithm for Neural Network Design and Application to Channel Equalization", *IMA Conference on Applications of Combinatorial Mathematics*, December 1994.

R.P. Lippmann, "An introduction to computing with neural nets," *IEEE ASSP Magazine*, Vol.4, pp.4-21, 1987.

W.S. McCulloch and W. Pitts, "A Logical Calculus of the Ideas Imminent in Nervous Activity", *Bulletin on Mathematical Biophysics* Vol. 5, pp. 115-133, 1943.

B. Mulgrew, K. Nisbet and S. McLaughlin, "Degeneracy of nonlinear predictor," in *Proc. VI European Signal Processing Conference*, EUSIPCO-92, Brussels, 1992, pp.791-794.

B. Mulgrew, E.S. Warner and P.M. Grant, "A Bayesian receiver for asynchronous CDMA communications", *in IEEE PIMRC Conference Proceedings* pp. 985-989, September 1995.

B. Mulgrew, "Applying Radial Basis Functions", *IEEE Signal Processing Magazine*, Vol. 13, No. 2, pp. 50-65, March 1996.

S.U.H. Qureshi, "Adaptive equalization," *Proc. IEEE*, Vol.73, No.9, pp.1349-1387, 1985.

D.E. Rumelhart, G.E. Hinton and R.J. Williams, "Learning internal representations by error propagation," in *Parallel Distributed Processing: Explorations in the Microstructure of Cognition*, D.E. Rumelhart and J.L. McClelland, Eds., Cambridge, MA: MIT Press, 1986, pp.318-362.

I. Tabus, M. Gabbouj and L. Lin, "Real domain adaptive WOS filtering using neural network approximations", *IEEE Winter Workshop on Nonlinear DSP*, pp. 7.2-1.1 to 7.2-1.6, Jan 1993.

H. Tong and K.S. Lim, "Threshold autoregression, limit cycles and cyclical data", *Journal Royal Statistical Society*, B Vol. 42, pp. 245-292, 1980.

A.S. Weigend, B.A. Huberman and D.E. Rumelhart, "Predicting the future: a connectionist approach", *Int. Journal Neural Systems*, Vol. 1, No. 3, pp. 193-209, 1990.

L. Yin, J. Astola and Y. Neuvo, "Adaptive stack filtering with application to image processing", *IEEE Trans. Signal Processing*, Vol. 41, No. 1, pp. 162-184, Jan 1993.