Nonlinear classification and adaptive structures

Cowan, Colin, Grant, Peter, Chen, Shang-Liang, Gibson, Gavin

Colin F. N. Cowan, Peter M. Grant, Shang-Liang Chen, Gavin J. Gibson, "Nonlinear classification and adaptive structures," Proc. SPIE 1348, Advanced Signal Processing Algorithms, Architectures, and Implementations, (1 November 1990); doi: 10.1117/12.23465

Event: 34th Annual International Technical Symposium on Optical and Optoelectronic Applied Science and Engineering, 1990, San Diego, CA, United States

Non-linear Classification and Adaptive Structures

C. F. N. Cowan, P. M. Grant, S. Chen and G. J. Gibson

Dept. of Electrical Engineering, University of Edinburgh,

Mayfield Road, Edinburgh EH9 3JL, Scotland.

ABSTRACT

The main purpose of this paper is to examine a number of possible architectures for non-linear adaptive filtering, specifically related to adaptive equalisation. The approach taken proceeds by first reformulating the filtering process as a form of classification task in N dimensions. In the case of filtering the dimensionality is determined by the number of data samples in the filter data input vector. The task of classification then proceeds using a number of possible strategies, i.e. the multilayer perceptron, Volterra series modeling and cluster analysis. The techniques are evaluated in comparison with normal linear equalisation procedures.

1. INTRODUCTION

In many cases the task of data equalisation in communications systems is viewed as an inverse filtering problem [1] in which the task of the equaliser is to form an approximation to the inverse of the distorting channel (which is linear). Thus the classical equaliser may be viewed as an inherently linear process and it might be assumed that no benefit could acrue from the introduction of non-linear processes.

However, if we restate the problem in a different way then this distinction is less clear, and indeed proves to be invalid. The assertion made here is that the task of the equaliser is to reconstruct the data sequence which was input to the communications channel, as accurately as possible. It must be remembered here that for a digital communications scenario that this sequence will consist of information chosen from a finite alphabet. In the situation where no additive noise exists this strategy will lead to an optimal equaliser which is indeed the linear inverse filter specified above. However, if there is any additive noise in the channel or the channel response is non-minimum phase then equalisation becomes a non-linear problem [2].

This may be best illustrated by examining a simple example in which we have a random binary input sequence {-1, 1} which is input to a channel with a minimum phase transfer function given by:

$$H(z) = 1 + 0.5z^{-1} \tag{1}$$

The equaliser to be used will form its estimate of the current data sample by operating on only the last 2 observations of the channel output:

$$y(n) = \sum_{i=0}^{\infty} h(n)\, s(n-i) \tag{2}$$

$$\hat{s}(n) = f\{y(n), y(n-1)\} \tag{3}$$

where $y(n)$ is the channel output at time n, $s(n)$ is the data input at n and $\hat{s}(n)$ is the estimate of $s(n)$ formed by the equaliser. The function $f\{\bullet\}$ is as yet undefined. For the example quoted we may tabulate all possible channel output pairs $\{y(n), y(n-1)\}$ as shown in table I. These may then be plotted on a 2-dimensional graph with axes $y(n)$ and $y(n-1)$ as shown in figure 1.

s(n)	s(n-1)	s(n-2)	y(n)	y(n-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 I: Set all possible channel output pairs for the channel $1 + 0.5\, z^{-1}$

Figure 1 shows each of the points plotted as either a X or a ●. The X means that this data pair represents an input, $s(n)$, of -1 and the ● represents an input at time n of $+1$. It can now be easily seen how the equalisation problem is viewed as one of the classification, i.e. the task is to assign regions within this observation space to represent inputs of either $+1$ or -1.

2. THE CLASSIFICATION PROBLEM

In looking at the classification problem we will continue to use the example quoted in the last section. Clearly, in the absence of noise this is very simple since we only need to draw a straight line (the line (a) in figure 1) and classify all points to the right of this line as representing an input of $+1$ and to the left, -1. This corresponds to the linear combiner, $f\{\bullet\}$ is linear convolution followed by the signum function.

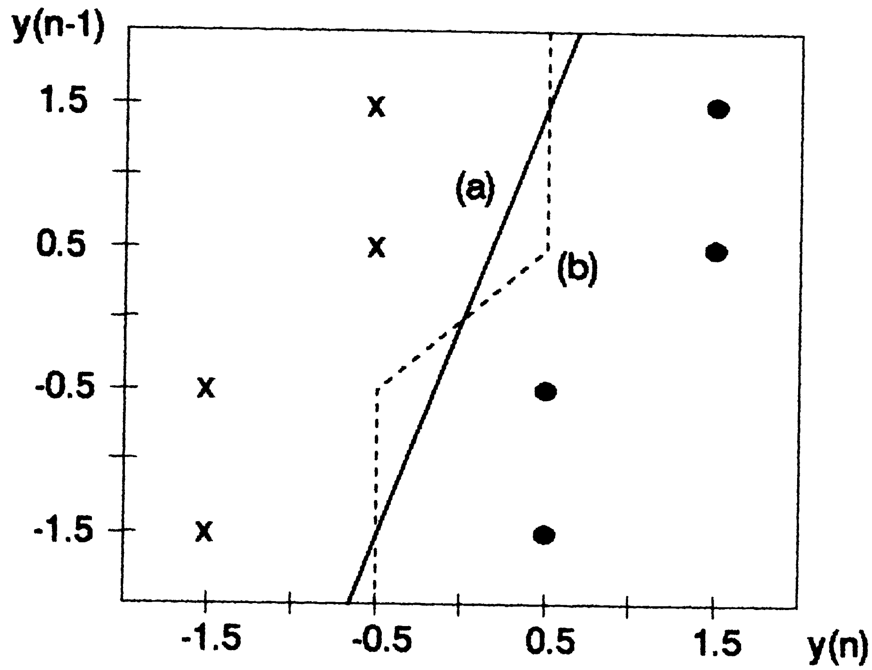

Figure 1: 2-dimensional representation of the symbol alphabet formed by 2 successive received samples from channel with impulse response $1 + 0.5z^{-1}$.

However, if we now add some noise to the observed channel outputs we will have some probability distribution function instead of the discrete points indicated in figure 1. If we then concentrate on the data points at ($-$ 0.5, 0.5) and (1.5, 0.5) we find that the -1 data point is closer to our linear decision boundary (indicated by the solid line (a) in Figure 1) than the +1 data point. Therefore, in the situation of additive noise there is a higher probability of the point at ($-$ 0.5, 1.5) being incorrectly detected as +1 than of the point at (1.5, 0.5) being incorrectly detected as $-$ 1. This is clearly a non-optimum situation.

If we apply a minimum a-posteriori error criterion to this problem then the optimum boundary becomes that shown by the broken line, (b) in figure 1. This non-linear boundary may not be formed by a simple linear combiner and fundamentally requires that $f\{\bullet\}$ in equation (1) is a non-linear operator [3].

3. THE MULTILAYER PERCEPTRON

In order to validate the non-linear classification process we first apply a very general non-linear operator to replace the linear combiner in the original equaliser. The structure chosen to do this is the so-called multilayer perceptron (MLP) [4]. Each perceptron consists of a linear combiner followed by a non-linear sigmoid function (figure 2). The overall equaliser is built up by a layered set of these giving 7 nodes in the 1st layer, 3 in the 2nd and one output node, figure 3. The use of 3 layers is adopted as this provides the best flexibility in terms of definition of the decision boundary [5]. The actual choice of number of nodes is rather difficult to justify and was in fact arrived at experimentally in this case.

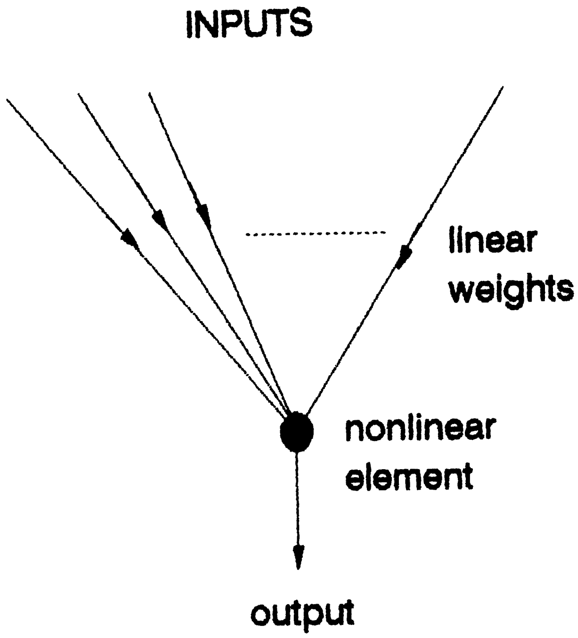

Figure 2: Schematic diagram of a single perceptron.

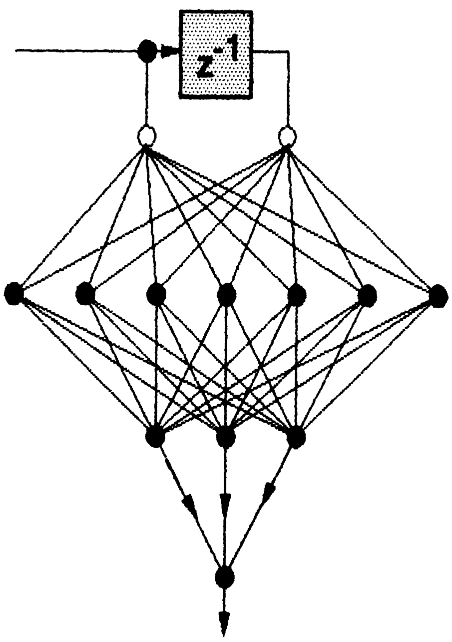

Figure 3: Schematic diagram of the complete equaliser using the multilayer perceptron structure.

For the example quoted before such an MLP was used in a simulation programme to equalise the channel defined in equation (1). The decision boundary achieved after convergence of the learning phase is shown in figure 4. This proves to be close to the expected boundary defined in figure 1.

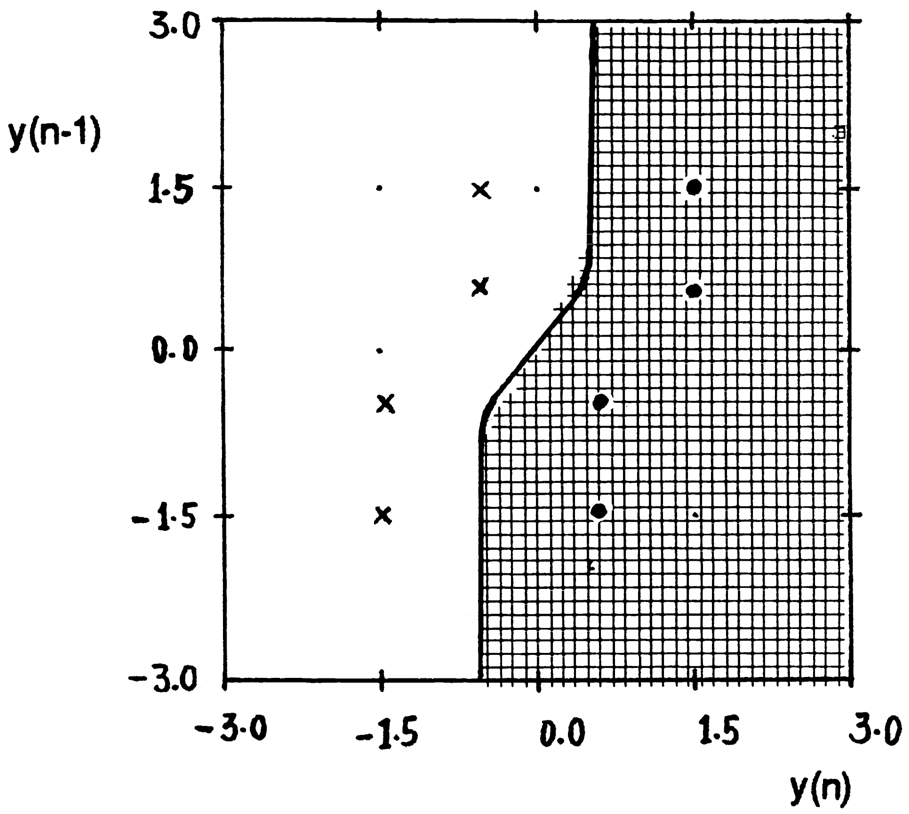

Figure 4: Converged decision boundaries for the MLP structure of fig. 3 using the channel defined in fig. 1.

The system was then evaluated to determine the bit error rate (BER) achieved by the classifier over a range of signal to noise ratios. The same test was applied to the linear equaliser system. These results are reproduced in figure 5 for sampled channel impulse response $0.3842 + 0.8704z^{-1} + 0.3842z^{-2}$ from which it can be seen that at very low noise levels there is little difference in performance but as the additive noise increases the MLP performs significantly better. This is consistent with the argument presented in section 2.

Comparisons have also been carried out when a decision feedback [1] structure is applied as an equaliser. Here, it was shown [6] that the performance enhancement persists and, further, that the MLP structure has a lesser sensitivity to error extension than the normal decision feedback equaliser.

Although this serves to illustrate the validity of the non-linear classification process in equalisation it does not lead to a practical structure from the implementation viewpoint.

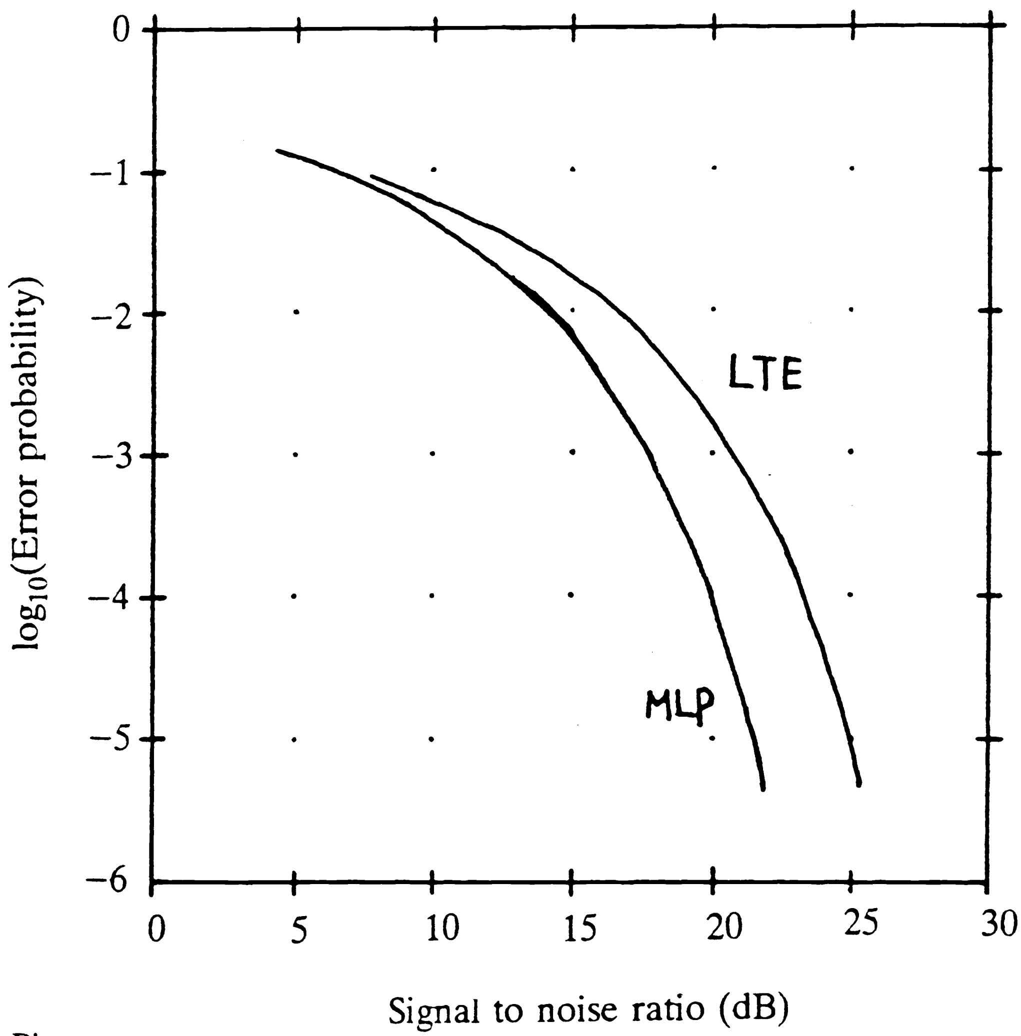

Figure 5: Bit error rate plotted against signal to noise ratio for the linear equaliser and the MLP equaliser.

The MLP structure is not only costly (in computational terms) but also cannot be guaranteed to converge using the back propagation algorithm [7] employed in these tests.

4. THE VOLTERRA MODEL

The fundamental problem affecting convergence of the MLP structure is the highly nested nature of the non-linear elements within the structure. This means that the adaptive coefficients may not be linearly related to the filter output resulting in highly irregular error surfaces containing multiple minima. One way of tackling this problem is to shift the (fixed) system non-linear operation to the input of the filter thus maintaining a linear relationship between the adaptive weights and the output error. Such a strategy will not be wholly effective due to the distortion of input statistics produced by the non-linear operators. There will also be problems introduced by large increases in dynamic range within this filter.

Perhaps the best known technique for implementing a structure of this type is the Volterra series [8] which provides an expansion of the following form:

$$\hat{s}(n) = a_o + \underline{H}_1 [\underline{y}(n)] + \underline{H}_2 [\underline{y}(n)] + \cdots$$

$$\text{where } \underline{H}_k(\underline{y}(n)) = \sum^{\tau_1} \sum^{\tau_2} .. \sum^{\tau_k} a_k (\tau_1, \tau_2 .. \tau_k)\, y(n-\tau_1) y(n-\tau_2) .. y(n-\tau_k) \qquad (4)$$

In such a system all the adaptive coefficients, a_n, are linear with respect to the output and thus should be more controllable. A structure of this type was applied to the same equalisation problem as before and the resulting classification boundary is shown in figure 6. Although the boundary here is perhaps not quite as well defined as the MLP example it should be noted that convergence was always achieved by this structure (convergence is also an order of magnitude faster than the equivalent MLP). However, it required a Volterra series of order 5 to achieve this result which does not really improve on the computational complexity of the MLP.

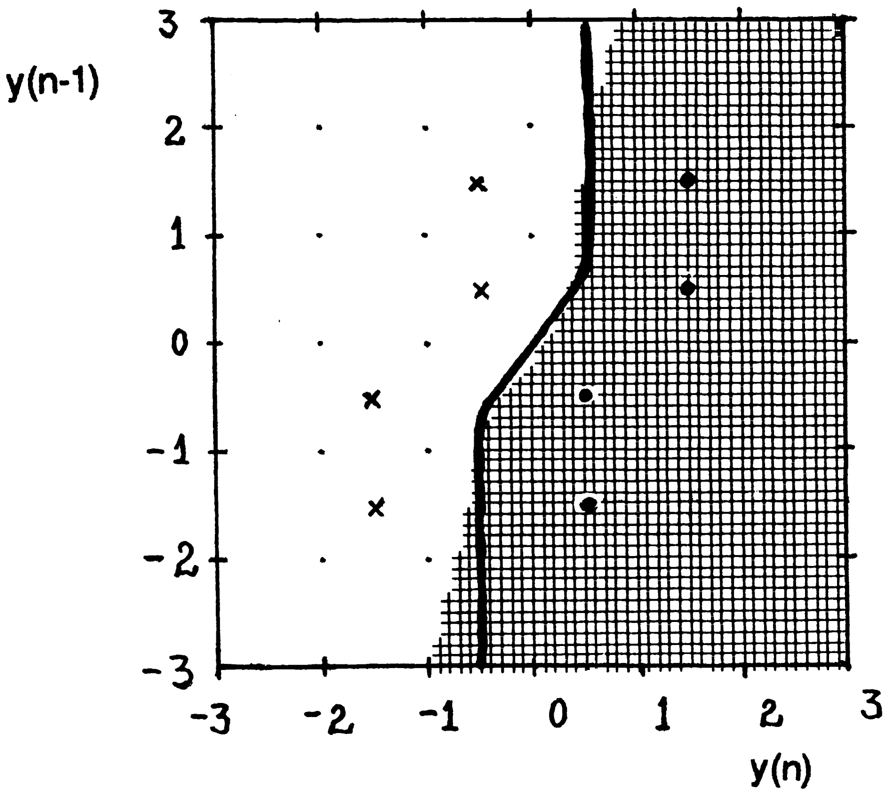

Figure 6: **Converged decision boundaries for the 5th order Volterra Series equaliser using the channel in fig. 1.**

The results in figure 7 for the same channel as figure 5 show the BER comparison between the MLP and Volterra equalisers. Clearly the MLP marginally out-performs the Volterra system. This is due to the tighter approximation involved in the latter case while the MLP is overdefined in terms of the non-linearity. It is this same effect which causes the fundamental differences in convergence characterisation.

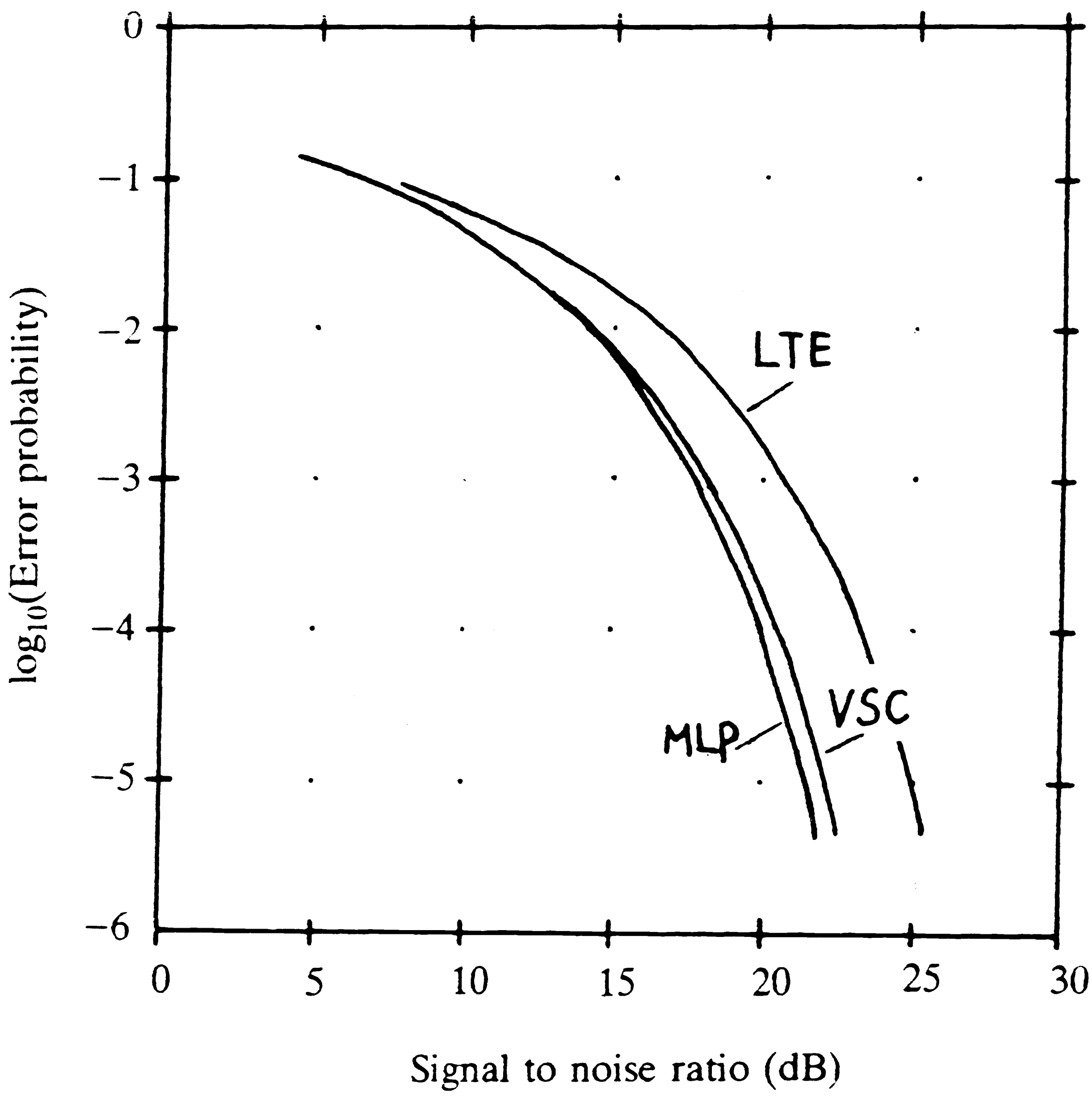

Figure 7: Bit error rate plotted against signal to noise ratio for the MLP equaliser and the Volterra Series equaliser.

5. CLUSTERING

The two preceeding examples show first how effective non-linear classification can be and second, how difficult it is to implement such a structure in practice. In this section we set out to determine how the problem may be further refined to yield a more appropriate practical equaliser.

One way of refining the classification structure is to reformulate the rationale by abandoning the idea of forming a decision boundary. In this case we would determine the correct outcome by calculating a statistical distance from each of the known outcomes. For the example taken we are using 2 output observations which, if there was no noise and no distortion, would result in 4 distinct observation outcomes, figure 8. Two of these outcomes correspond to a -1 and the other two to a $+1$. If these points are taken as our initial estimates of cluster centres we may proceed to build up a cluster model using the so-called Mahanalobis distance [9] classifier as follows:

$$d_i = (\underline{y}(n) - <\underline{y}_i(n)>)^t \, \underline{r}_i^{-1}(n) \, (\underline{y}(n) - <\underline{y}_i(n)>)$$

$$<\underline{y}_i(n+1)> = \lambda <\underline{y}_i(n)> + \underline{y}(n)$$

$$\underline{r}_i^{-1}(n+1) = \frac{1}{\lambda}\left\{\underline{r}_i^{-1}(n) - \frac{\underline{r}_i^{-1}(n)\,\underline{y}(n)\,\underline{y}^t(n)\,\underline{r}_i^{-1}(n)}{\lambda + \underline{y}^t(n)\underline{r}_i^{-1}(n)\underline{y}(n)}\right\}$$

$$0 < \lambda \leq 1 \qquad (5)$$

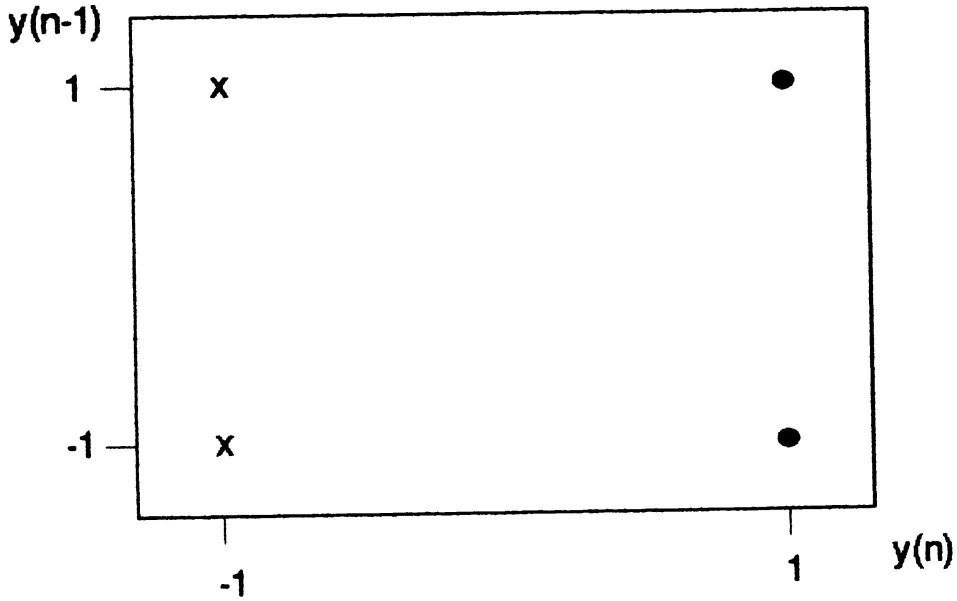

Figure 8: 2-dimensional representation of the received signal alphabet formed by 2 successive received samples for a non-distorting channel.

This strategy will allow the cluster shape to evolve to match the split and spread of the observation points evidenced in figure 1 but avoids the explicit use of a decision boundary. Further, the use of the recursions in (5) ensures optimal convergence rates (this should be compared with the recursions used in recursive least squares adaptive algorithms [10]). The exponential window implied by the use of λ here provides an ability to cope with nonstationary situations.

An interesting further point regarding this strategy is that since we are explicitly identifying input signal pairs which must occur in overlapped sequence we will have explicit information on the possible transitions between points from one sample to the next. These possible transitions are illustrated in figure 9. This information could be used to further refine our output estimate by using a maximum liklihood decoding structure.

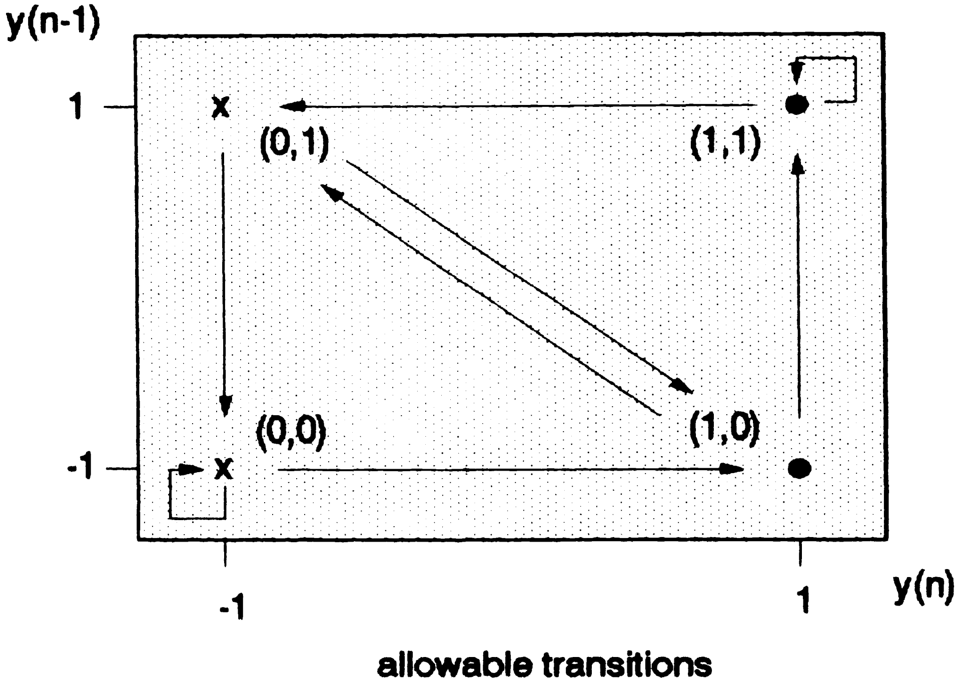

Figure 9: Possible sample by sample transitions between symbols for the situation in fig. 8.

6. CONCLUSIONS

It has been shown that non-linear processing has certain fundamental advantages in the equalisation of digital communications channels in comparison to their linear counterparts. Two such non-linear discriminators have been introduced, the multilayer perceptron and the Volterra series. The MLP was shown to have superior discrimination power but at the expense of poorer convergence properties. Neither structure is attractive from the viewpoint of computational complexity.

The final section introduced an architecture which could, potentially, provide a solution to these problems in a robust and efficient manner. The authors have not yet done a practical evaluation on this structure but would expect the system to at least equal the performance of the Volterra system. The added enhancement of the maximum likelihood decoder should enhance achievable bit error rate considerably.

ACKNOWLEDGEMENT

We acknowledge the generous sponsorship of these studies by the UK Science & Engineering Research Council under award GR/E/10357.

7. REFERENCES

[1] Qureshi, S.U.H.; "Adaptive equalisation", IEEE Comm. Soc. Mag, Vol. 21, No. 2, pp. 9-16, March 1982.

[2] Cowan, C.F.N., Gibson, G.J. and Siu, S.; "Data equalisation using highly non-linear adaptive architectures", Proc. SPIE Conf. on 'Advanced algorithms and architectures for signal processing IV', San Diego, Aug. 1989, pp. 34-43.

[3] Gibson, G.J., Siu, S. and Cowan, C.F.N.; "Multilayer perceptron structures applied to adaptive equalisers", Proceedings IEEE Int. Conf. on ASSP, 1989, pp. 1183-1186.

[4] Widrow, B., Winter, R.G. and Baxter, R.A.; "Layered neural nets for pattern recognition", IEEE Trans. ASSP, vol. 36, no. 7, July 1988, pp. 1109-1118.

[5] Gibson, G.J. and Cowan, C.F.N.; "On the decision regions of multilayer perceptrons", Proc. IEEE, to be published, 1990.

[6] Siu, S., Gibson, G.J. and Cowan, C.F.N.; "Decision feedback equalisation using neural network architectures", IEE Int. Conf. on Artificial Neural Networks, IEE Conf. Pub. no. 313, London, 1989, pp. 125-128.

[7] Rumelhart, D.E., Hinton, G.E. and Williams, R.J.; "Learning internal representations by error propagation", in Rumelhart and McClelland (eds.), 'Parallel distributed processing: explorations in the microstructure of cognition', MIT Press, 1986.

[8] Chen, S., Gibson, G.J. and Cowan, C.F.N.; "Adaptive channel equalisation using a polynomial-perceptron structure", IEE Proc. Pt. I, to be published, 1990.

[9] Mirzai, A.R., Cowan, C.F.N. and Crawford, T.M.; "Intelligent alignment of waveguide filters using a machine learning approach", IEEE Trans. Microwave Theory and Tech., vol. 37, no. 1, January 1989, pp. 166-173.

[10] Cowan, C.F.N.; "Performance comparisons of finite linear adaptive filters", IEE Proc. Pt. F., vol. 134, no. 3, June 1987, pp. 211-216.