ADAPTIVE BAYESIAN DECISION FEEDBACK EQUALISER BASED ON A RADIAL BASIS FUNCTION NETWORK

S. Chen, B. Mulgrew and S. McLaughlin
Department of Electrical Engineering, University of Edinburgh,
King's Buildings, Edinburgh EH9 3JL, U.K.

Abstract

The paper derives a novel Bayesian decision feedback equaliser (DFE) for digital communications channel equalisation. It is shown how decision feedback is utilized to improve equaliser performance as well as to reduce computational complexity. The relationship between the Bayesian solution and the radial basis function (RBF) network is emphasized and two adaptive schemes are described for implementing the Bayesian DFE using the RBF network. The maximum likelihood sequence estimator (MLSE) and the conventional DFE are used as two bechmarks to assess the performance of the Bayesian DFE.

1. Introduction

Adaptive equalisation is an important technique for combating distortion and interference in communications links. There are basically two categories of equalisers, namely the sequence-estimation and symbol-decision equalisers. The optimal sequence-estimation equaliser is the MLSE [1], which is optimal for detecting the entire transmitted sequence and provides the best attainable performance for any equalisers. High complexity and deferring decisions are however two drawbacks of the MLSE. Symbol-decision equalisers are more commonly seen and are typically based on adaptive linear filter design, e.g., the conventional DFE [2]. The linear filter approach has a very simple computational requirement but does not achieve the optimal solution for the symbol-decision structure.

In the past few years, considerable advance has been achieved in adaptive nonlinear equaliser design based on artificial neural networks [3]-[5]. The attraction of neural network equalisers is their ability to adaptively form the optimal Bayesian solution for the symbol-decision structure and therefore to provide significant performance gain over the conventional linear filter approach. Sections of the communications community however are sceptical of this emerging nonlinear technique. This is largely because a nonlinear adaptive filter often results in a substantial increase in computational complexity. The main purpose of the present study is to demonstrate that this need not be the case for adaptive nonlinear equaliser design. By an intelligent use of decision feedback, the computational complexity of the adaptive Bayesian equaliser can be reduced to a level close to the conventional DFE in general or even less in a special case.

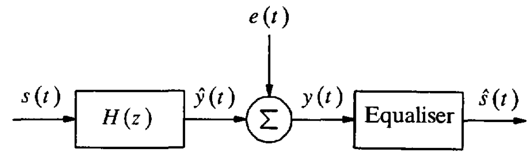

Fig.1.Discrete-Time Model of Data Transmission System.

The digital communications scenario is illustrated in the baseband discrete time model depicted in Fig.1, where a digital sequence $s(t)$ is transmitted through a dispersive channel with transfer function

$$H(z)=\sum_{i=0}^{n} h_i z^{-i}. \qquad (1)$$

$s(t)$ is assumed to be an equiprobable and independent binary sequence taking values from $\{\pm 1\}$. The channel output is corrupted by an additive white Gaussian noise $e(t)$, and $s(t)$ and $e(t)$ are assumed to be uncorrelated. The structure of a generic symbol-decision equaliser with decision feedback is shown in Fig.2, where τ is known as the equaliser delay, m and k are referred to as the feedforward and feedback orders respectively. The present study derives the optimal solution for the equaliser structure of Fig.2 based on the Bayes decision theory [6]. It is shown that the Bayesian solution has an

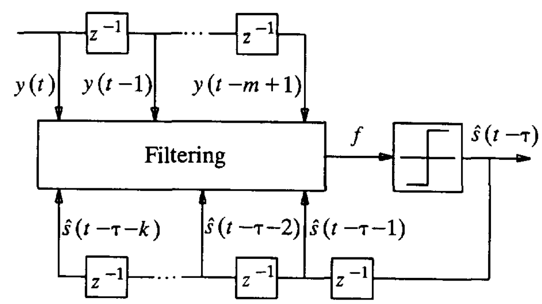

Fig.2.Schematic of Symbol-Decision Equaliser with Decision Feedback.

identical structure to the RBF network [7]. Two adaptive approaches are developed for realizing the Bayesian equaliser using the RBF network and the computational complexity of this adaptive nonlinear equaliser is compared with that of the conventional DFE. The performance of the Bayesian DFE is also compared with those of the MLSE and the conventional DFE. The 2-ary PAM signaling scheme is assumed in the present study simply for maximum clarification of the basic concepts. All the results in this paper can easily be extended to the general PAM scheme [8] and the QAM scheme with a complex-valued channel [9].

2. Bayesian decision feedback equaliser

For the channel (1) and an equaliser feedforward order m, there are $n_s=2^{n+m}$ combinations of the channel input sequence $\mathbf{s}(t)=[s(t)\cdots s(t-m+1-n)]^T$. Thus the noise-free channel output vector $\hat{\mathbf{y}}(t)=[\hat{y}(t)\cdots \hat{y}(t-m+1)]^T$ has n_s desired states. It is sufficient to consider a feedback order $k=m+n-1-\tau$. Assume that the correct feedback vector $\mathbf{s}_f(t-\tau)=[s(t-\tau-1)\cdots s(t-\tau-k)]^T$ is used and denote $n_f=2^k$ combinations of $\mathbf{s}_f(t-\tau)$ as $\mathbf{s}_{f,i}$, $1\le i\le n_f$. The set of desired channel states $Y_{m,\tau}$ can be divided into n_f subsets conditioned on $\mathbf{s}_f(t-\tau)$:

$$Y_{m,\tau}=\bigcup_{1\le i\le n_f} Y_{m,\tau,i}, \tag{2}$$

where $Y_{m,\tau,i}=Y^+_{m,\tau,i}\cup Y^-_{m,\tau,i}$, $Y^+_{m,\tau,i}=\{\hat{\mathbf{y}}(t)\,|\,s(t-\tau)=1\cup \mathbf{s}_f(t-\tau)=\mathbf{s}_{f,i}\}$ and $Y^-_{m,\tau,i}=\{\hat{\mathbf{y}}(t)\,|\,s(t-\tau)=-1\cup \mathbf{s}_f(t-\tau)=\mathbf{s}_{f,i}\}$. It is straightforward to verify that the optimal Bayesian equaliser conditioned on $\mathbf{s}_f(t-\tau)=\mathbf{s}_{f,i}$ is defined as:

$$\hat{s}(t-\tau)=\begin{cases}1, & f_B(\mathbf{y}(t)\,|\,\mathbf{s}_f(t-\tau)=\mathbf{s}_{f,i})\ge 0,\\ -1, & f_B(\mathbf{y}(t)\,|\,\mathbf{s}_f(t-\tau)=\mathbf{s}_{f,i})<0,\end{cases} \tag{3}$$

with the optimal conditional decision function given by

$$f_B(\mathbf{y}(t)\,|\,\mathbf{s}_f(t-\tau)=\mathbf{s}_{f,i})=\sum\alpha\exp(-\|\mathbf{y}(t)-\mathbf{y}_l^+\|^2/2\sigma_e^2) -\sum\alpha\exp(-\|\mathbf{y}(t)-\mathbf{y}_l^-\|^2/2\sigma_e^2), \tag{4}$$

where $\mathbf{y}(t)=[y(t)\cdots y(t-m+1)]^T$, the two sums are over $\mathbf{y}_l^+\in Y^+_{m,\tau,i}$ and $\mathbf{y}_l^-\in Y^-_{m,\tau,i}$ respectively, σ_e^2 is the noise variance, and α is an arbitrary positive constant.

It is interesting to note that the feedback vector is only used to narrow down the number of desired states that have to be considered at each t. As a result, only $2^{\tau+1}$ states are required to compute the Bayesian decision function at each t. Without decision feedback, all the n_s states would be required. Furthermore it can be shown that $m=\tau+1$ is sufficient for the Bayesian DFE [9]. That is, a Bayesian DFE of $m=\tau+1$ has the same performance as those of $m>\tau+1$. Because the equaliser order is generally $m>\tau+1$ without decision feedback, the reduction factor in computational complexity owing to decision feedback is larger than 2^n. In general, the minimum distance between $\mathbf{y}_l^+$ and $\mathbf{y}_l^-$ considered at each t is increased due to decision feedback and this is why decision feedback improves performance.

Equaliser delay $\tau=0$ is a particularly interesting case. In this case $m=1$ is sufficient for the Bayesian DFE and only the current observation $y(t)$ is used as the feedforward input. The conditional Bayesian decision function takes a very simple form:

$$f_B(y(t)\,|\,\mathbf{s}_f(t)=\mathbf{s}_{f,i})=(y(t)-y_i^-)^2-(y(t)-y_i^+)^2. \tag{5}$$

Computational load for this Bayesian DFE is surely less than what required for a conventional DFE. When the channel is normalized, theoretical error probability of this Bayesian DFE under the assumption of correct bits feedback can be shown to be

$$Q(|h_0|/\sigma_e)=\int_{|h_0|/\sigma_e}^{\infty}(2\pi)^{-1/2}\exp(-y^2/2)dy, \tag{6}$$

comparing to $Q(1/\sigma_e)$ for the ideal channel without ISI.

For the Bayesian DFE of $\tau>0$, (6) does not apply but this result does suggest a general rule in choosing equaliser delay for the Bayesian DFE so as to minimize the achievable error probability in symbol detection. If h_j is the channel tap that has the largest magnitude, equaliser delay should be chosen as $\tau=j$. The other two structure parameters can then be set to $m=\tau+1$ and $k=n$.

3. Adaptive implementation and complexity

The Bayesian DFE can conveniently be implemented using the RBF network. The RBF network [7] is a two-layer processing structure depicted in Fig.3. The hidden layer consists of some computing nodes and the output layer is a linear combiner. The network response is given by:

$$f_r(\mathbf{y})=\sum_{i=1}^{n_h} w_i\,\phi(\|\mathbf{y}-\mathbf{c}_i\|^2/\rho_i), \tag{7}$$

where n_h is the number of hidden nodes, $\mathbf{c}_i$ are the RBF centres, ρ_i are the widths of the nodes and w_i are the weights. The network (7) obviously has an identical form to the Bayesian filter (4). n_h can be set to n_s. The hidden nodes are grouped in accordance with the conditional subsets $Y_{m,\tau,i}$ and the centres $\mathbf{c}_j$ realize corresponding channel states $\mathbf{y}_j\in Y_{m,\tau,i}$. The feedback vector determines which subset of hidden nodes should be active at t. The function ϕ is selected as $\phi(y)=\exp(-y)$ and all the widths are chosen to be $\rho_i=2\sigma_e^2$. The weights are set to either 1 or -1 correspondingly and the RBF network will realize precisely the Bayesian DFE.

Adaptively adjusting the RBF centres so that they converge to the channel states is the key and two alternative schemes are available. The first method estimates the channel model based on the least mean square (LMS) algorithm and uses the channel estimate to calculate a subset of centres. The second method estimates the channel states directly based on a clustering algorithm [5]. During the training period, the following supervised clustering algorithm can be used

$$\begin{aligned}&\text{if } (\mathbf{s}(t) == \mathbf{s}_j)\ \{\\ &\quad \mathbf{c}_j(t) = counter_j * \mathbf{c}_j(t-1) + \mathbf{y}(t);\\ &\quad counter_j \mathrel{+}= 1;\\ &\quad \mathbf{c}_j(t) = \mathbf{c}_j(t)/counter_j;\ \}\end{aligned} \tag{8}$$

For nonstationary channels, the following adaptive version of (8) is preferred

$$\text{if } (\mathbf{s}(t)==\mathbf{s}_j)\ \mathbf{c}_j(t)=\mathbf{c}_j(t-1)+g_c*(\mathbf{y}(t)-\mathbf{c}_j(t-1)); \tag{9}$$

where g_c is a learning rate. During data transmission a decision-directed version of (9) can be employed [5].

The computational loads of the Bayesian DFE based on the LMS channel estimator and the clustering algorithm (9) are listed in the Table, where the complexity of the conventional DFE with the LMS adaptation is also given. Consider a 4-tap ($n=3$) channel and first assume that h_1 has the largest magnitude. As mentioned early, $\tau=1$ can be chosen for the Bayesian DFE. This Bayesian DFE based on the LMS estimator needs 4 exp(·)s, 32 multiplications and 36 additions while 4 exp(·)s, 14 multiplications and 19 additions are required if adaptation is done using the clustering algorithm. For the conventional DFE with a structure of $\tau=1$, $m=5$ and $k=6$, 23 multiplications and 22 additions are required. If the magnitude of h_0 is the largest and $\tau=0$ is chosen, 16 multiplications and 15 additions are required for the Bayesian DFE using the LMS estimator while only 3 multiplications and 5 additions are needed based on the clustering algorithm.

4. Simulation results

The channel $H(z)=0.7255+0.5804z^{-1}+0.3627z^{-2}+0.0724z^{-3}$ was used to illustrate the two adaptive schemes. For this example, $\tau=0$ and $m=1$ were chosen. The SNR was set to 15dB and adaptive gain in the LMS channel estimator was 0.05. The estimated states obtained using the channel estimator and the clustering algorithm (8) are shown in Figs. 4 and 5 respectively, where the dotted lines indicate the desired states. Because the number of the channel taps is much smaller than that of the channel states, the channel-estimator approach requires a far shorter training sequence than the clustering approach. The first approach updates every centre at each t while only one centre is changed

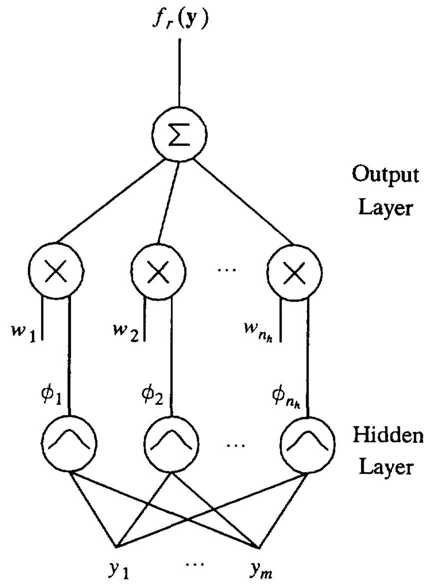

Fig.3.Schematic of Radial Basis Function Network.

Bayesian DFE $\tau>0$ with LMS	
exp(·)s:	$2^{\tau+1}$
multiplications:	$2\times n+3+2^{\tau+1}\times(\tau+2)+\sum_{i=0}^{\tau}\{(n-\tau+i)+2^{\tau+1}\times(\tau+1-i)\}$†
additions:	$2\times n+2^{\tau+2}\times(\tau+1)+1+\sum_{i=0}^{\tau}\{(n-\tau-1+i)+2^{\tau+1}\times(\tau+1-i)\}$†
Bayesian DFE $\tau>0$ with clustering	
exp(·)s:	$2^{\tau+1}$
multiplications:	$2^{\tau+1}\times(\tau+2)+\tau+1$
additions:	$2^{\tau+2}\times(\tau+1)+2\times\tau+1$
Bayesian DFE $\tau=0$ with LMS	
multiplications:	$3\times n+7$
additions:	$3\times n+6$
Bayesian DFE $\tau=0$ with clustering	
multiplications:	3
additions:	5
Conventional DFE with LMS	
multiplications:	$2\times(m+k)+1$
additions:	$2\times(m+k)$

Table. Comparison of Computational Complexity.
† Estimated upper bound for computing subset states

at each t in the second approach. The former is therefore better suited for highly time-varying channels. The latter however has a much simpler computational requirement. It does not rely on the linear channel assumption and is immune from nonlinear channel distortion.

The bit error rate of the Bayesian DFE was compared with those of the conventional DFE and the MLSE using the channel $H(z)=0.3482+0.8704z^{-1}+0.3482z^{-2}$. $\tau=1$ was chosen for the conventional and Bayesian DFEs. The Bayesian DFE had $m=2$ and $k=2$ while the conventional DFE was defined by $m=4$ and $k=4$. The coefficients of the conventional DFE were set to the Wiener solution. The MLSE was implemented as a Viterbi algorithm with the exact channel model and a fixed decision delay. Fig.6 shows the performance of the conventional and Bayesian DFEs, where it is seen that with detected bits feedback and at error probability 10^{-4} the Bayesian DFE had more than 2dB improvement in SNR over the conventional DFE. From Fig.6, it seemed that error propagation was not serious for the Bayesian DFE and it also appeared that the Bayesian DFE was less affected by error propagation compared to the conventional DFE. The three performance curves shown in Fig.7 were obtained by the Bayesian DFE with detected bits feedback and the MLSE with fixed decision delays 1 and 10 respectively. The performance of the Bayesian DFE appeared to be compatible to that of the MLSE with the same decision delay 1. The MLSE achieved better results with higher decision delay and increased computational complexity.

5. Conclusions

A Bayesian solution has been derived for digital communications channel equalisation with decision feedback. A novel strategy of utilizing decision feedback has been proposed which not only improves equaliser performance but also reduces computational complexity dramatically. It has been shown that the Bayesian solution has an identical structure to the RBF network, and two adaptive approaches have been developed to realize the Bayesian DFE using the RBF network. The scheme based on a channel estimator requires a shorter training sequence while the clustering scheme offers lower computational load and greater immunity to nonlinear channel distortion. The MLSE and the conventional DFE have been used as two bechmarks to assess this Bayesian DFE.

6. Acknowledgments

This work was supported by SERC and DTI of UK. The authors acknowledge stimulating discussions with Professor P.M. Grant on the topics reported in this study.

7. References

[1] 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.

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

[3] 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.

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

[5] S. Chen and B. Mulgrew, "Overcoming co-channel interference using an adaptive radial basis function equaliser," *EURASIP Signal Processing J.,* Vol.28, No.1, 1992.

[6] R.O. Duda and P.E. Hart, *Pattern Classification and Scene Analysis.* New York: John Wiley and Sons, 1973.

[7] S. Chen, C.F.N. Cowan and P.M. Grant, "Orthogonal least squares learning algorithm for radial basis function networks," *IEEE Trans. Neural Networks,* Vol.2, No.2, pp.302-309, 1991.

[8] S. Chen, P.M. Grant and C.F.N. Cowan, "Orthogonal least squares algorithm for training multi-output radial basis function networks," submitted to *IEE Proc. Part F,* Special Issue on Artificial Neural Networks, 1992.

[9] S. Chen, S. McLaughlin and B. Mulgrew, "Complex-valued radial basis function network, Part II: application to digital communications channel equalisation," submitted to *EURASIP Signal Processing J.,* 1992.

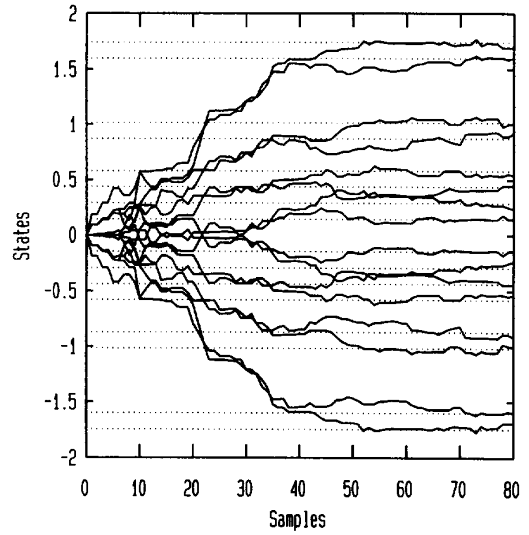

Fig.4.Trajectories of Estimated Channel States Based on the Channel-Estimator Approach.

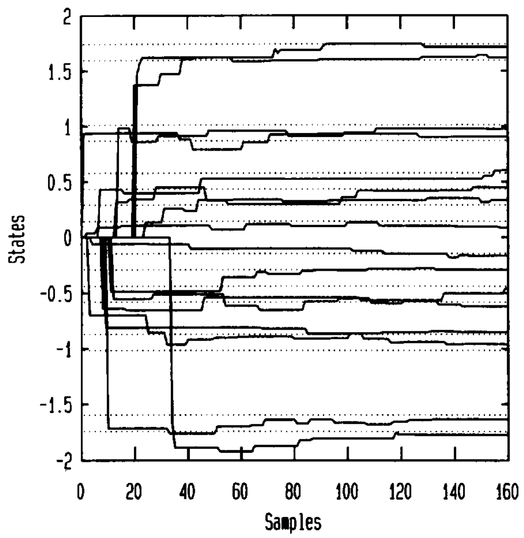

Fig.5.Trajectories of Estimated Channel States Based on the Clustering Approach.

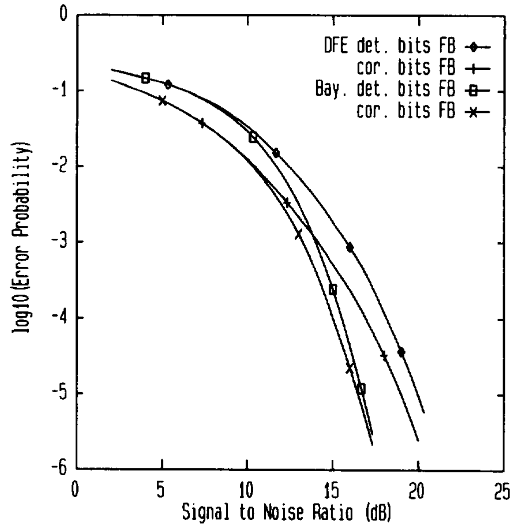

Fig.6.Performance Comparison. Bayesian and conventional DFEs.

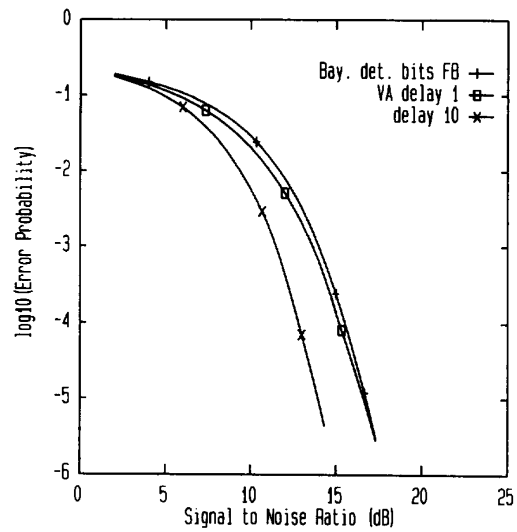

Fig.7.Performance Comparison. Bayesian DFE and Viterbi algorithm.