EFFICIENT CLUSTERING TECHNIQUES FOR SUPERVISED AND BLIND CHANNEL EQUALIZATION IN HOSTILE ENVIRONMENTS

Sergios Theodoridis and Kristina Georgoulakis
University of Athens
Department of Informatics, TYPA Buildings, 15771 Athens Greece
Tel : + (301) 7211119 - Fax : + (301) 7228981
e-mail: stheodor@di.uoa.gr, kristina@di.uoa.gr

ABSTRACT

In this paper the equalization problem is treated as a classification task. No specific (linear or nonlinear) model is required for the channel or for the interference and the noise. Training is achieved via a supervised learning scheme. Adopting Mahalanobis distance as an appropriate distance metric, decisions are made on the basis of minimum distance path. The proposed equalizer operates on a sequence mode and implements the Viterbi searching Algorithm. The robust performance of the equalizer is demonstrated for a hostile environment in the presence of CCI and non linearities, and it is compared against the performance of the MLSE and a symbol by symbol RBF equalizer. Suboptimal techniques with reduced complexity are discussed. The operation of the proposed equalizer in a blind mode is also considered.

1 INTRODUCTION

Intersymbol Interference (ISI) is a major impairment in todays high bit rate Communications Systems. Channel equalizers, used in the receiver part, aim to suppress the effect of ISI [1]. The presence of channel nonlinearities as well as Co - Channel Interference (CCI) further degrade systems' performance. Equalization under such hostile environments is a difficult task, which channel equalizers have to cope with.

Maximum Likelihood Sequence Estimation (MLSE) is a robust way to combat ISI leading to enhanced performance compared to symbol by symbol equalizers (i.e., Linear Transversal Equalizer (LTE), Decision Feedback Equalizer (DFE), Radial Basis Function Equalizer (RBF)). However, MLSE performance is seriously degraded in the presence of CCI [2]. The reason is twofold. CCI is neither Gaussian nor white. Thus the Euclidean metric used in the classical Maximum Likelihood (ML) equalizers no more approximates reality. Furthermore, the channel estimates using standard Least Square Techniques are no more BLUE (Best Linear Unbiased Estimates) in the presence of non white interference [3]. Moreover, when nonlinearities are present, non linear modeling of the channel is required, which for severe nonlinearities is not always a straightforward task.

All the equalizers mentioned in the previous paragraph adopt specific models for the channel, the noise and the interference - either by explicitly estimating the channel (MLSE) or by modeling the decision boundary as a specific function (symbol by symbol equalizers). A different view to the equalization was adopted in [4]. According to this approach the equalizer is treated as a classifier, thus freeing itself from the need of an explicit adoption of specific models both for channel and interferences. This equalizer is named Clustering Based Sequence Equalizer (CBSE) and is described in paragraph 3.

In CBSE (as in every Viterbi type equalizer) the complexity is exponentially dependent on the channel length [5]. Thus, for long channels, CBSE complexity can be very high. In paragraph 4 three techniques are described for CBSE complexity reduction. In the first technique, a suboptimal methodology, called "Selection of Clusters through Per Survivor Processing", is proposed in order to reduce complexity. Complexity reduction of the second method is achieved through the use of M-Algorithm. The third one is based on grouping the clusters together, resulting in a smaller number of clusters.

Furthermore, in paragraph 5 performance results of CBSE in a hostile environment are described. CBSE performance is compared to the performance of the MLSE as well as the recently suggested method of [2]. The results demonstrate the robustness of the proposed approach with respect to the nature of interference, noise and channel. Moreover, simulations show that with the incorporation of the three reduction techniques mentioned above, low complexity is achieved with only little sacrifice in the CBSE performance.

Finally, the issue of operation of CBSE in blind mode is studied in paragraph 6. The operation of CBSE in blind mode exploits the fact that all information, needed to perform the equalization task, is hidden in the structure of the received data and in the allowable transitions between them.

2 SYSTEM DESCRIPTION

Consider the signal $s(t)$ at the output of the channel :

$$s(t) = \sum_{i=0}^{n} h(i) I(t-i) \quad (1)$$

where $I(t)$ is an equipropable sequence of bipolar data (± 1) and $h(i)$ is the impulse response of the channel. When nonlinearities exist in the channel the transmitted signal takes the form of a polynomial :

$$r(t) = \sum \lambda_i (s(t))^i, i = 0, 1, ... \quad (2)$$

The received signal is :

$$g(t) = r(t) + w(t) + c(t) \quad (3)$$

where $w(t)$ is the noise sequence and $c(t)$ the Co Channel Interference component, given by:

$$c(t) = \sum_{j=1}^{k} \sum_{i=0}^{n_j} h_j(i) J_j(t-i) \quad (4)$$

with k being the number of interferers and $J_j(i)$ the sequences of bipolar data applied to the interfering sources. Symbols $I(t)$ and $J_j(t)$ are considered to be statistically independent. In the following k is assumed to be equal to 1, a valid assumption for a mature telecommunications system [6].

The general model of a data transmission system in the presence of noise, CCI and nonlinearities is shown in Figure 1. The values of the gain coefficients $D_1, ..., D_{n-1}$ determine the severity of the nonlinear distortion.

The Signal to Noise Ratio (SNR) and the Signal to Interference Ratio (SIR) are determined as:

$$SNR = \frac{\sigma_s^2}{\sigma_e^2} \qquad SIR = \frac{\sigma_s^2}{\sigma_c^2} \tag{5}$$

with σ_e^2 the variance of the noise, $\sigma_s^2 = E[s^2(k)]$ and $\sigma_c^2 = E[c^2(k)]$, where E[·] is the expectation operator.

3 CBSE

According to this approach no specific modeling is required both for the channel as well as the interferences. That is, instead of trying to adopt a specific model for the decision boundary (equivalently the channel, the noise and the interference) the method focuses on the clusters, which the received data form. The received data are clustered around specific points, whose number and constellation shape is determined by the spread of the channel and the impairments characteristics. Let us consider, for simplicity, two successively received samples, in the absence of any distortion, as a point in the two dimensional space. The randomness of noise leads to the formation of a cluster around this point, formed by the possible positions of these two symbols. The variance of the noise determines the radius of the cluster. The existence of ISI causes a movement and an increase of the number of clusters. Specificly, if n is the number of symbols over which ISI is spread, then the number of clusters is multiplied by 2^n. CCI causes a further movement and increase of the number of clusters (for each of the clusters we have 2^{2+n_1} new ones, where $n_1 + 1$ is the number of taps of the interfering channel) [2]. When on top of the previous impairments there is also nonlinear distortion, each of the clusters moves to a new position, depending on the form of the nonlinearities.

The CBSE scheme stems from the above observations. Let us assume that ISI spreads over n symbols and received samples are treated, for simplicity, in groups of two. We denote by $\mathbf{g} = [g(t), g(t-1)]$ the vector of received data and $\mathbf{c}_i = [c_i(t), c_i(t-1)]$ the vectors of clusters' centers (which correspond to the noise free outcomes of the channel) for $i = 1, 2, ..., N$. If the span of the channel over the transmitted symbols I(t) is $n+1$ then

$$\begin{aligned} c_i(t) &= f[I(t), \ldots, I(t-n)], \\ c_i(t-1) &= f[I(t-1), \ldots, I(t-n-1)] \end{aligned} \tag{6}$$

where $i = 1, \ldots, N$ and f is a nonlinear mapping. It is obvious from (6) that clusters, where successive samples reside, are not independent, due to the n common transmitted symbols shared between two successive samples. Hence only specific transitions among the different clusters are possible. Thus a Viterbi type procedure for a minimum path search can be constructed, with $I(t-1), \ldots, I(t-n-1)$ being the states, provided a distance metric is adopted.

A popular distance metric in classification problems is the Mahalanobis distance defined as

$$D_i = (\mathbf{g} - \mathbf{c}_i)^T \Sigma_i^{-1} (\mathbf{g} - \mathbf{c}_i) \tag{7}$$

where Σ_i is the covariance matrix of each cluster defined as

$$\Sigma_i = E[(\mathbf{g} - \mathbf{c}_i)(\mathbf{g} - \mathbf{c}_i)^T] \tag{8}$$

The non diagonal choice of Σ_i takes care of the underlying correlation in the presence of non white interference [4]. Furthermore, when for reasons of reducing complexity or even when the exact spread of the channel is unknown, the number of adopted clusters is smaller than the true one, then a grouping of clusters takes place. Thus the resulting clusters have no more spherical distribution, even in the case of white Gaussian noise. So, the use of non diagonal Σ_i permits the exploitation of the underlying shape of the clusters [7].

Training of the centers ($\mathbf{c}_i$) is equivalent to label them as +1 or -1 and it is achieved during the training period. Matrix Σ_i can be similarly estimated and adapted. In the sequel we have assumed that the covariance matrix is independent on the specific cluster, that is : $\Sigma_i \equiv \Sigma, \forall i$. Adaptation is also possible during the decision directed mode [4].

4 COMPLEXITY REDUCTION TECHNIQUES

The major drawback of sequence equalizers is their complexity, due to the exponential dependence on the channel spread. Thus for binary data and a channel length of $n+1$ the complexity of CBSE is of order 2^{n+1}, which is the number of states in CBSE. For channels with long impulse response the complexity of CBSE becomes too high. In the following, three methods for reducing the complexity of CBSE are described. All these techniques are suboptimal and computational savings are obtained by reducing the number of states, through which the trellis searching is evolved.

The technique of Per Survivor Processing has been successfully used in modulated data transmission systems subject to ISI and in systems which employ Trellis Coded Modulation, for complexity reduction [8]. The idea of Per Survivor Processing is adopted here for CBSE complexity reduction. In particular, it is suggested that fewer states are used than the number of clusters suggests. That is instead of searching among all possible clusters (2^{n+2}), we reduce the searching to a smaller group of clusters (2^{n+2-m}), by assuming that a number m of symbols in the history of each surviving path is correctly identified. In other words we use the history of each surviving path to make up for the reduced number of states assumed. Let us take for example, a channel of length $n+1 = 5$, leading to 64 clusters in the two dimensional space. We assume a trellis structure with 8 states (instead of 32). Thus each trellis branch is related to 4 transmitted data $(I(t), I(t-1), I(t-2), I(t-3))$ but there is no information about the two remaining data $(I(t-4), I(t-5))$ which are also needed to uniquely define a cluster in the received signal constellation. Compensation for the lack of these two unknown information bits comes from the survivor path corresponding to that branch.

In the second method the M-Algorithm is used for complexity reduction. Basically, the M-Algorithm is a modified Viterbi Algorithm, which instead of searching the full trellis and keeping one survivor path per state, it keeps only the M best surviving paths [9]. Thus, the complexity of M-Algorithm is related to the number M of surviving paths, and therefore , for $M < 2^{n+1}$ (2^{n+1} the number of states), a significant reduction in complexity is obtained. The same idea has been adopted in our case for the CBSE technique.

The third method for reducing complexity of CBSE has been exploited in [10] and has already been mentioned in paragraph 3. According to this method instead of assuming the true number of clusters ($= 2^{n+2}$ for the two dimensional case) formed by the received samples, we assume fewer number of clusters : $2^g < 2^{n+2}$. Thus, cluster grouping takes place. In this way complexity of CBSE reduces to 2^{g-1} , which is the new number of states, or equivalently the number of survivor paths (one survivor path to each state).

5 PERFORMANCE RESULTS

A channel widely used in the literature has been adopted for the simulations :

$$H(z) = 0.3482 + 0.8704z^{-1} + 0.3482z^{-2} \quad (9)$$

The interfering channel is assumed to be [2]

$$H(z) = 0.6 + 0.8z^{-1} \quad (10)$$

The SIR is chosen low in order to account for the augmented levels of CCI in modern communication systems [6].

The receivers studied are : a)The new Clustering Based Sequence Equalizer, b)The conventional MLSE implemented with the Viterbi Algorithm and c)The RBF based equalizer described in [2]. The RBF and the proposed equalizer assume in every case 2^{2+n} cluster centers, or equivalently 2^{n+1} states. The corresponding states of the MLSE are 2^n. The function of channel nonlinearities used for the purposes of our simulation is [4], (fig.1)

$$r(t) = s(t) + 0.15(s(t))^2 + 0.01(s(t))^3 \quad (11)$$

Figure 2 presents the results from simulations on the assumed channel with nonlinearities and SIR =10db. From the figure, the robust performance of CBSE in such a hostile environment is verified. In contrast it is shown that MLSE and RBF performance is substantially degraded in the presence of CCI, and channel nonlinearities. It should be noted that the performance of MLSE can be improved by adopting nonlinear models for the channel. However, our aim here is to demonstrate that CBSE need not to bother about such a modeling and the whole procedure is the same independent of the presence or not of nonlinearities. In the same figure appears the effect of clusters grouping on the performance of CBSE.

For the same channel and impairments as above, Figure 3 summarizes the loss of performance versus complexity when Per Survivor Processing (PSP) is used with CBSE. From the figure it is apparent that the performance of CBSE with PSP (4 and 2 states) is close to the optimum (c1) and is substantially better compared to the performance of CBSE with clusters grouping for the same number of states.

Figure 4 highlights the results of implementation of M-Algorithm to CBSE. From these curves is verified that the performance of CBSE with the M-Algorithm is very close to the optimum, even for a very low complexity (2,4 states).

The above results have also been verified for a number of different channels. All the results presented above demonstrate the enhanced performance possibilities offered by the CBSE. *The fact that no explicit assumptions about the noise, interference and channel are required, provides the scheme with extra degrees of freedom to be able to learn the environment and to perfom robust equalization. Moreover, the problem of CBSE complexity can efficiently be solved by means of the reduction complexity techniques examined.*

6 BLIND CBSE

In this section we approach blind equalization problem from the clustering point of view, in an unsupervised mode of operation. Since no training sequence is available, labeling of each of the clusters is not possible in a direct way. The method we propose here constitutes of the following steps :

a) The clusters are identified via an unsupervised clustering technique, i.e. isodata [7]. b) The received samples are grouped as feature vectors, two dimensional in our case, and are allocated to the clusters according to the minimum distance, based on an adopted distance measure, i.e. Euclidean or Mahalanobis. Due to the interdependence of successively received samples, only specific jumps, among clusters, can take place. Thus, a table is formed providing the information of the jumps which take place among the identified clusters. This information can be used for the identification of the clusters. Let us consider for example the simple channel

$$H(z) = 1 + 0.5z^{-1} \quad (12)$$

If we group together two successively received samples (g(t),g(t-1)), then the possible jumps from the cluster corresponding, say, to (I(t-2),I(t-1),I(t)) = (1,1,1) can only be to clusters corresponding to (1,1,1) and (1,1,-1). Figure 5 demonstrates the possible jumps in a parent-child tree dependency. A table, which records the possible jumps for each one of the clusters, is constructed by recording the jumps over a period of observations. For example, we can record that once in cluster A, the next observation will cluster either to A or to B, and so on. Since errors are inevitable, the decision about the children clusters (to which jumps are made) of a parent cluster (from which jumps are made), is based on a winner criterion rule. In our case was the two most probable ones, that is the two clusters to which most of the jumps had been done from the parent cluster. The observation period is obviously problem dependent and the SNR is a major factor. For example, for the above mentioned two taps channel, we found that 80 observations were enough to reveal the parent-child pattern for an SNR=10db.

Having constructed the parent-child table for each of the clusters, labeling for each of them can take place. First, we identify the one (of the two) clusters which is a child of itself (i.e. jumps to itself), say cluster A. This must be either (+1 + 1 +1) or (-1 -1 -1). Assuming that it is (-1 -1 -1), then B (its other child) is (-1 -1 +1) (figure 5). Then exploiting the parent-child relationships and the fact that the other cluster that jumps to itself (H) (fig.5) will necessarily be the (+1 +1 +1), it is not difficult to see that labeling all the clusters in a tree is trivial. The ambiguity between (+1 +1 +1) or (-1 -1 -1) is not important, because it results to a mirror sequence and can be surmounted by Differential Coding [11]. Once labeling has been completed, a Viterbi type algorithm can be used.

References

[1] J.G.Proakis "Digital Communications" N.Y., Mc Graw Hill 1983

[2] S.Chen and B.Mulgrew "Overcoming CCI using an adaptive RBF equalizer" IEEE Trans. on Signal Processing Vol.28, 1992, pp.91-107

[3] N.Kalouptsidis, S.Theodoridis "Adaptive System Identification and Signal Processing Algorithms" Prentice Hall 1994

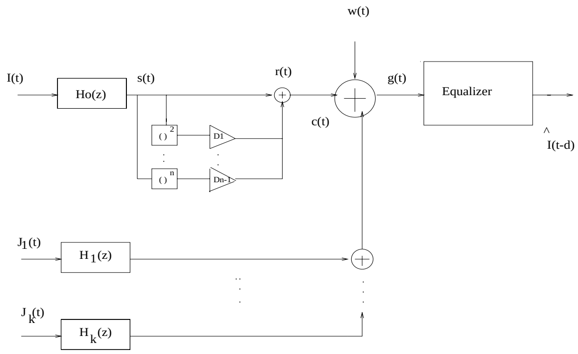

Figure 1: Data transmission system with co-channel interferers and nonlinear impairments.

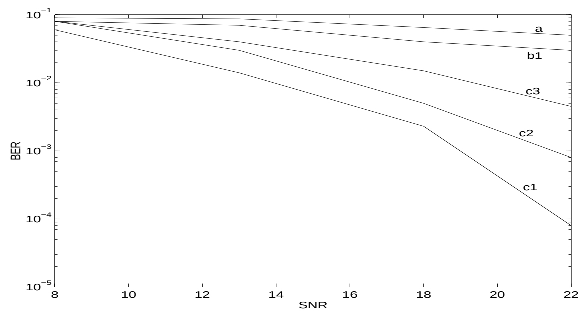

Figure 2: Channel $H(z) = 0.3482 + 0.8704z^{-1} + 0.3482z^{-2}$, interfering channel $H(z) = 0.6 + 0.8z^{-1}$, SIR=10db and nonlinear impairments. a:RBF-16 centers, b1:MLSE-4 states, c1:CBSE-8 states, c2:CBSE-4 states, c3:CBSE-2 states

[4] S.Theodoridis,C.F.N.Cowan,C.P.Callender and C.M.S.Lee "Schemes for equalization of Communications channels with nonlinear Impairments" IEE Proceedings Communications Vol 142, no 3, June 1995, pp.165-171

[5] G.D.Forney "The Viterbi ALgorithm" Proceeding of the IEEE Vol61 , No 3, March 1973, pp 268-278.

[6] N.W.K.Lo, D.D. Falconer, A.U.H.Sheikh "Adaptive equalizer for CCI in a Multipath Fading Environment" IEEE Trans. on Communications Vol 43, No2/3/4, 1995, pp.1441-53

[7] R.O.Duda and P.E.Hart "Pattern Classification and Scene Analysis" New York, J.Wiley 1973

[8] R.Raheli, A.Polydoros, C.K.Tzou "Per-Survivor Processing" Digital Signal Processing No3, 1993, pp175-87

[9] F.Jelinek and J.B.Anderson "Instrumentable tree encoding of information sources" IEEE Trans. on Information Theory, volIT-22, Jan-71, pp82-83

[10] C.F.N.Cowan "Communications Equalization Using Non-Linear Adaptive Networks", Conference Record on the 25th Asilomar Conference on Signals, Systems and Computers Vol.1, 1991 pp.199-202

[11] N.Seshadri "Joint data and channel estimation using blind trellis search techniques" IEEE Transactions on Communications, vol42, no2/3/4, 1994, pp.1000 - 1011

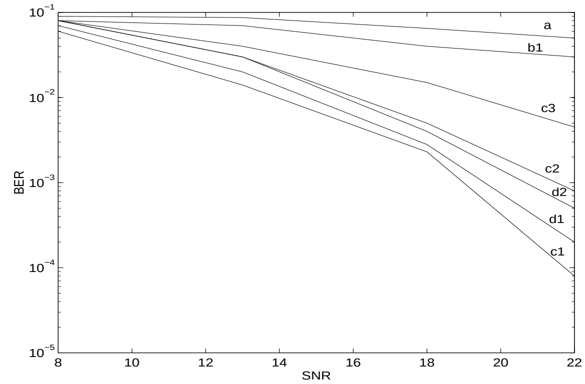

Figure 3: Channel $H(z) = 0.3482 + 0.8704z^{-1} + 0.3482z^{-2}$, interfering channel $H(z) = 0.6 + 0.8z^{-1}$, SIR=10db and nonlinear impairments.a:RBF-16 centers, b1:MLSE-4 states, c1:CBSE-8 states, c2:CBSE-4 states, c3:CBSE-2 states, d1:CBSE-PSP-4 states, d2:CBSE-PSP-2 states

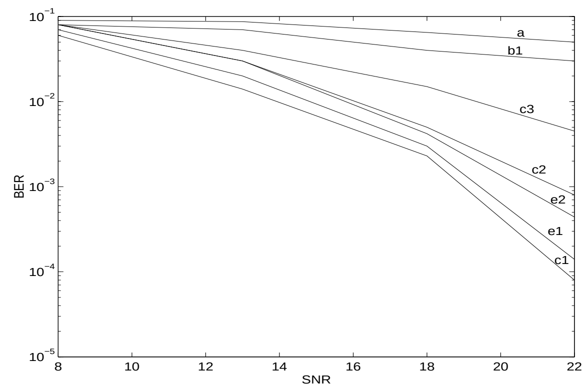

Figure 4: Channel $H(z) = 0.3482 + 0.8704z^{-1} + 0.3482z^{-2}$, interfering channel $H(z) = 0.6 + 0.8z^{-1}$, SIR=10db and nonlinear impairments.a:RBF-16 centers, b1:MLSE-4 states, c1:CBSE-8 states, c2:CBSE-4 states, c3:CBSE-2 states, e1:CBSE-M.A.-4 states, e2:CBSE-M.A.-2 states

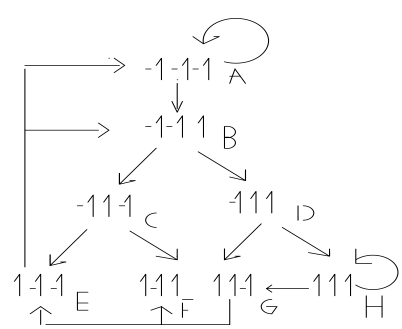

Figure 5: Label of Clusters and parent-child relationship.