A Neural Predictor for Blind Equalization of Digital Communication Systems

Charles C. Cavalcante*
Univ. Fed. do Ceará
Fortaleza CE, Brazil
charlesc@dee.ufc.br

Jugurta R. Montalvão F.*
Univ. Tiradentes
Aracaju SE, Brazil
Jugurta.Montalvao@int-evry.fr

Bernadette Dorizzi
*Inst. Nat. des Télécom.
Evry, France
Bernadette.Dorizzi@int-evry.fr

João Cesar M. Mota
Univ. Fed. do Ceará
Fortaleza CE, Brazil
mota@dee.ufc.br

Abstract

In digital channel equalization, self-learning techniques are used in the cases where a training period is not available. Considering the transmitted sequence as composed of independent random variables, the equalization task can be done by means of prediction. In this work we propose to use Artificial Neural Networks (ANN), instead of a linear prediction device, in order to obtain a better performance. Prediction concepts are revisited and a new self-organized algorithm is proposed to update the first layer in the nonlinear predictor whose aim is to avoid local minimum points in the applied cost function. The second layer is updated by using a classical supervised algorithm. Simulation results are presented which illustrate the performance of this technique.

1. Introduction

Equalization of digital communication channels is usually done by using a transmitted sequence also known to the receiver during a preamble period. Figure 1 depicts a simplified digital communication system.

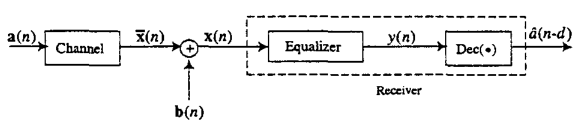

Figure 1: Digital Communication System

where $\mathbf{a}(n)$ is the transmitted sequence, $\mathbf{b}(n)$ is the noise sequence and $\hat{a}(n-d)$ is the estimated symbol after a delay d.

Self-learning (blind) equalizers are used in order to provide the correct identification of transmitted symbols when one does not have a training period or when it is not practical to use such a strategy, as in digital TV broadcasting and multipoint networks where training has to be redone whenever one single receiver is inserted in the system. Another example is mobile communication systems, where due to multipath fading, the received signal may be so low that the receiver does not synchronize adequately.

Some classical strategies for blind equalization are the following related algorithms: Direct Decision (DD), Sato, Godard [1], Benveniste-Goursat [2] and Shalvi-Weinstein [3].

Considering the transmitted symbols to be uncorrelated, it is possible to deal with the blind equalization problem by means of prediction [4]. In this context, the pioneer work is that of Macchi and Hachicha, in 1986, who used a linear filter as a predictor device. The symbol with the desired information is recovered, in this case, by elimination of the existing redundancy in the time sequence formed by the channel outputs.

In classical implementations for minimum phase channels, the prediction filter is linear and has a finite impulse response which is adapted to minimize the prediction squared error. This, indeed, is equivalent to a whitening process over the received time sequence and the white sequence obtained in the predictor error is the same as that of the transmitted symbols apart for a scale factor. The prediction error sequence will be i.i.d. if the transmitted sequence $\{\mathbf{a}(n)\}$ is also i.i.d and the noise negligeable.

Nevertheless, in spite of the well established theory behind linear prediction, a crucial point should be mentioned: if the channel is nonminimum phase, even an infinite length predictor cannot provide a super-whitened error sequence. In other words, the original transmitted sequence cannot be recovered as result of the intrinsic linear mapping of past samples on the current estimated one. Nonetheless, it is quite easy to show that, in most cases, the ideal mapping is nonlinear (see example in Section 2).

Therefore, in this work, we propose a nonlinear structure based on Artificial Neural Networks as a prediction device. Moreover, in order to improve the adaptive solution, we divided the learning task in to two steps. First a new self-organized learning algorithm is proposed to adapt the first layer then the second layer connections are updated by

0-7803-5800-7/00/$10.00 ©2000 IEEE

means of a classical supervised algorithm (supervised with respect to the prediction error, but blind with respect to the transmitted symbols).

In Section 2, we explore the prediction concepts. Section 3 is dedicated to the new proposed self-organized algorithm. In Section 4, some simulation results are presented to illustrate the performance of this new strategy and, in the last section, conclusions are presented.

2. Prediction Concepts

In digital communication systems, the implicit goal of applying prediction is to remove temporal redundancies in the received signal, which can be used in blind equalization. The representation of a prediction-based equalizer is shown in Figure 2, where $\mathbf{x}(n)$ is the noisy channel output sequence, $\hat{x}(n)$ is the predicted signal, $e(n)$ is the prediction error, $\mathbf{P}$ is a prediction filter and g is an Automatic Gain Control (AGC).

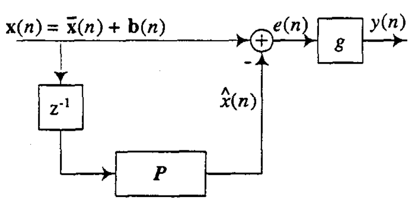

Figure 2: Prediction-Based Equalizer

The channel is modeled as a linear filter with discrete impulse response represented by

$$F(z) = \sum_{i=0}^{N-1} f_i z^{-i} \tag{1}$$

where f_i are the channel coefficients and N is the channel length. We can also represent the channel model in a vectorial form: $\mathbf{f} = [\, f_0 \; f_1 \; f_2 \; \cdots \; f_{N-1} \,]^T$.

Therefore, the noiseless channel outputs, which we call *channel states*, can be written as:

$$\begin{aligned} \bar{x}(n) &= a(n)f_0 + \cdots + a(n-N+1)f_{N-1} \\ \bar{x}(n-1) &= a(n-1)f_0 + \cdots + a(n-N+2)f_{N-1} \\ \bar{x}(n-2) &= a(n-2)f_0 + \cdots + a(n-N+3)f_{N-1} \\ \vdots \quad & \quad \vdots \qquad\qquad \vdots \qquad\qquad \vdots \end{aligned} \tag{2}$$

And then, the prediction error corresponds to:

$$e(n) = x(n) - \mathbf{P}(\mathbf{x}(n-1)) \tag{3}$$

where $\mathbf{x}(n-1) = [\, x(n-1) \; x(n-2) \; \cdots \,]^T$, $x(n) = \bar{x}(n) + b(n)$ and $\mathbf{P}$ is a function which provides a prediction of $x(n)$.

The simplest structure is obtained by choosing $\mathbf{P}$ to be linear. Although it is the most frequently used function, the application of the corresponding linear predictor is limited to minimum and maximum phase channels [1, 4]. Some works (for instance [4] and references therein) have proposed a combined structure to treat nonminimum phase channels.

To remove temporal redundancies, the prediction error equation is rewritten in the form:

$$\begin{aligned} e(n) &= a(n)f_0 + a(n-1)f_1 + \cdots + b(n) \\ &-\mathbf{P}\left(\bar{x}(n-1) + b(n-1) + \bar{x}(n-2) + b(n-2) + \cdots\right) \end{aligned} \tag{4}$$

Using a linear filter with discrete finite impulse response $\mathbf{p} = [\, p_1 \; p_2 \; p_3 \cdots \; p_k]$, as a predictor device, we have:

$$\begin{aligned} e(n) &= \underbrace{a(n)f_0 + a(n-1)f_1 + \cdots + b(n)}_{x(n)} \\ &- \underbrace{[x(n-1)p_1 + x(n-2)p_2 + \cdots + x(n-k)p_k]}_{\hat{x}(n)} \end{aligned} \tag{5}$$

where p_i is the i-th prediction filter coefficient.

Expanding $\hat{x}(n)$ leads to:

$$\begin{aligned} \hat{x}(n) &= \left(a(n-1)f_0 + a(n-2)f_1 + \cdots + b(n-1)\right)p_1 \\ &+ \left(a(n-2)f_0 + a(n-3)f_1 + \cdots + b(n-2)\right)p_2 + \cdots \\ &+ \left(a(n-k)f_0 + a(n-k+1)f_1 + \cdots + b(n-k)\right)p_k \end{aligned} \tag{6}$$

Combining Equations (5) and (6) leads to:

$$\begin{aligned} e(n) &= a(n)f_0 + b(n) + a(n-1)\left[f_1 - f_0p_1\right] \\ &-b(n-1)p_1 + a(n-2)\left[f_2 - f_1p_1 - f_0p_2\right] - b(n-2)p_2 \\ &+ \cdots - a(n-N+1)f_{N-1}\left[p_k\right] - b(n-N+1)p_k \end{aligned} \tag{7}$$

The goal here is to recover $a(n)f_0$. For this purpose, we must remove the undesired symbols by adapting the prediction filter in order to force them to zero. Unfortunately, not all coefficients can be canceled at once.

It becomes evident that, there is a *residue* in the prediction error expression and this residue cannot be cancelled by a finite linear filter. For equalization to be achieved, the samples of the prediction error sequence must to be uncorrelated and this residue must also be negligeable with respect to $a(n)f_0$. A possible solution to this problem is to increase the predictor order which decreases the contribution of the residue. The scale factor multiplying $a(n)$ is recovered by the AGC that matches the power of sequence $\mathrm{e}(n)$ and transmitted sequence $\mathbf{a}(n)$.

However it is known that in the nonminimum phase channel case, it does not work and in any case the noise itself cannot be removed.

Since the linear mapping of a linear predictor may be not enough for equalization, we have tryed to find a structure able to perform a nonlinear mapping in a satisfactory way. We chose the function implemented by an ANN that has the following form:

$$\mathbf{F} = \sum_i \beta_i \cdot \text{sign}\,(x - \theta_i) \tag{8}$$

where sign$(\cdot)$ is the signum function.

In the nonlinear case, Equation (3) is rewritten by replacing function $\mathbf{P}$ by a nonlinear function $\mathbf{F}_{NN}$ where the subscript stands for a neural network.

$$e(n) = x(n) - \mathbf{F}_{NN}\,(\mathbf{x}(n-1)) \tag{9}$$

According to Equation (8), $\mathbf{F}_{NN}$ is a sum of weighted and shifted copies of sign$(\cdot)$, whose parameters β and θ would be found by means of an a priori knowledge of the channel coefficients. However, since we do not have such an a priori knowledge, all parameters of $\mathbf{F}_{NN}$ are stochastically adjusted by means of the new algorithm described in Section 3.

Expanding Equation (9), it follows that:

$$\begin{aligned} e(n) &= a(n)f_0 + a(n-1)f_1 + \cdots + b(n) \\ &-\mathbf{F}_{NN}\,(x(n-1)\,,\, x(n-2)\,,\,\cdots) \end{aligned} \tag{10}$$

It is possible to find a function $\mathbf{F}_{NN}$ such that we can exactly cancel the term: $a(n-1)f_1 + \cdots + a(n-N+1)f_{N-1}$. Moreover this function can only explicitly depend on $x(n-1)$ since it has all information about past symbols that we need. So, rewriting Equation (10) it follows that:

$$\begin{aligned} e(n) = a(n)f_0 + a(n-1)f_1 + \cdots + b(n) - \\ \underbrace{\mathbf{F}_{NN}\,(a(n-1)f_0 + a(n-2)f_1 + \cdots + b(n-1))}_{a(n-1)f_1 + a(n-2)f_2 + \cdots + a(n-N+1)f_{N-1}} \end{aligned} \tag{11}$$

In this case we can obtain no residue. It is worth noting that the noise $b(n)$ is assumed to be an white Gaussian random variable and, consequently, it is not predictable, therefore the best the ANN can do is to cancel redundancies in the time sequence $\mathrm{e}(n)$.

Figure 3 shows, a two-dimensional illustration of a nonlinear mapping done by the ANN.

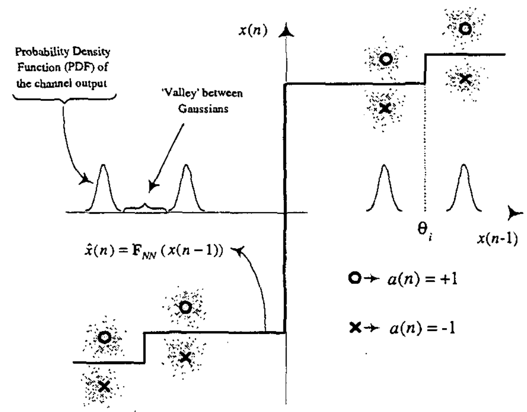

Figure 3: Nonlinear Mapping Function

Clearly, the parameters θ_i, in Equation (8), have a crucial role on the construction of ψ_{NN}. So, the problem of finding good parameters for the ANN is addressed in Section 3.

3. New Self-Organized Learning Algorithm

In classical techniques, such as backpropagation [5], the updating of parameters θ_i and of the synaptic weights is usually done through the same procedure. However, in this equalization problem such a procedure may not be able to quickly realize the fast transitions shown in Figure 3[1].

In order to solve this problem in a satisfactory way, we propose a new self-organized learning algorithm that is based on the minimization of a cost function in order to correctly find the θ_i of the neurons and to simplify the task of finding the interpolation surface.

It is easy to show that the probability density function (PDF) of the received signal is a mixture of Gaussians. Furthermore, the variance of each Gaussian is the noise variance and their means are channel-characteristic-related.

It can be seen from Figure 3 that the function referred to in Equation (8) can achieve correct interpolation if parameters θ_i, associated to the step transitions, are well placed between the 'valleys'. Then, we minimize a cost function that permits us to find those parameters by lookin for a function that can fit in those 'valleys'. Since the valleys have the

[1] In a previous work we have used the backpropagation algorithm for the equalization of channels with low intersymbol interference (ISI) [6].

shape of a "V", perhaps the simplest function, similar to a "V" we can use is $N(x,\theta_i) = |x-\theta_i| + \kappa$ where $\kappa > 0$ (see Figure 4). We use this function in order to simplify the resulting algorithm.

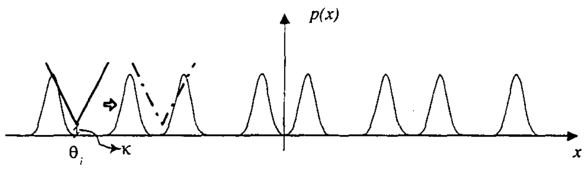

Figure 4: Looking for the 'valleys'

The constant κ is inserted to avoid instability problems when $|x-\theta_i|$ is very small and to guarantee a strictly positive function $N(x,\theta_i)$.

In order to measure function similarities, we have chosen the Kullback-Leibler Divergence (KLD), which is indeed a distance measure in the Riemann space [5] given by:

$$D_{h(x)||g(x)} = \int_{-\infty}^{\infty} h(x) \ln\left(\frac{h(x)}{g(x)}\right) dx \qquad (12)$$

where $h(x)$ and $g(x)$ are two strictly positive functions. Since the PDF of $x(n-1)$ and $N(x,\theta_i)$ are strictly positives, we apply KLD in order to measure similarities between them.

Eliminating the term which does not depend on θ_i we obtain a cost function given by:

$$J(\theta_i) = \int_{-\infty}^{\infty} p(x).\ln\left(\frac{1}{N(x,\theta_i)}\right) dx$$
$$J(\theta_i) = -\mathbb{E}\{\ln(N(x,\theta_i))\} \qquad (13)$$

In order to find the minima of this cost function we have to set $\frac{\partial J(\theta_i)}{\partial \theta_i} = 0$, where

$$\frac{\partial J(\theta_i)}{\partial \theta_i} = -\mathbb{E}\left\{\frac{\text{sign}(x-\theta_i)}{|x-\theta_i|+\kappa}\right\} \qquad (14)$$

The minimum value of θ_i can thus be obtained by using a simple stochastic version of the gradient algorithm:

$$\theta_i(n+1) = \theta_i(n) - \lambda \frac{\text{sign}(x(n-1)-\theta_i)}{|x(n-1)-\theta_i|+\kappa} \qquad (15)$$

which is a local adaptation rule of the Anti-Hebbian kind [5].

Finally, this adaptation rule is applied on the first layer while the second one is updated by a stochastic LMS algorithm.

4. Simulation Results

Performance measures of the proposed nonlinear predictor are compared with those of the linear predictor using two kinds of channels, firstly the minimum phase, where the linear predictor works, and then the nonminimum phase, where it gives bad results.

Using BPSK modulation and an SNR defined as $\text{SNR} = 10\log_{10}\left(\frac{\sigma_a^2 \sum_{i=0}^{N-1} f_i^2 + \sigma_b^2}{\sigma_b^2}\right)$ where σ_a^2 and σ_b^2 are the symbol and noise variance respectively, we simulated a 100 Monte-Carlo trials the Decision Squared Error (DSE) $(\varepsilon(n) = y - \text{Dec}(y))^2$ performance and the measure of Symbol Error Ratio (SER) for both cases.

Minimum Phase Channel

The minimum phase channel used in the computational simulations has the following impulse response:

$$f(z) = 1 + 0.8z^{-1} + 0.4z^{-2}$$

In the linear predictor we used a filter with 25 coefficients, a step factor equal to 10^{-3} and the initialization is done by setting the vector of filter coefficients at zero.

The nonlinear predictor has one input, 15 neurons in the hidden layer and one output. For this structure we used the following parameters: supervised learning rate equals 10^{-3}, λ equals 5.10^{-4}. The number of symbols for finding the θ_i was set to 50 and $\kappa = 10^{-7}$. The algorithm for the AGC [4] has a step size equal to 10^{-3} and the weights in the output layer were initialized at zero and the θ_i were randomly initialized from an uniformly distributed interval: [-1.5,1.5].

We can see in Figures 5 and 6 the performance achieved when the SNR is equal to 40 dB and 30 dB, respectively. Figure 7 shows the Symbol Error Rate (SER) for some SNRs computed for both predictors.

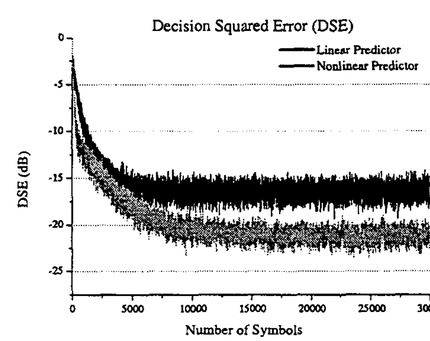

Figure 5: Decision-Squared Error (SNR = 40 dB) Minimum Phase Channel

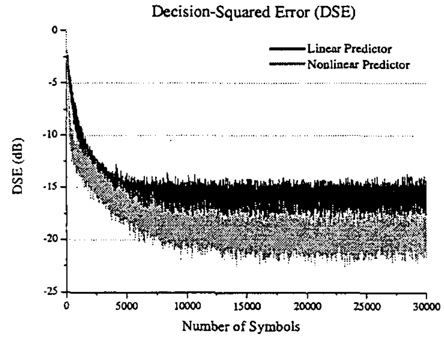

Figure 6: Decision-Squared Error (SNR = 30 dB)
Minimum Phase Channel

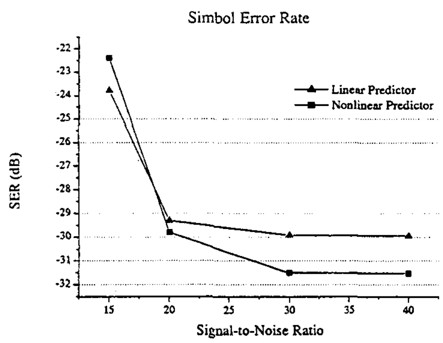

Figure 7: Symbol Error Rate
Minimum Phase Channel

One can easily see that the performance of the nonlinear predictor is better than that of the linear one when SNR > 20 dB.

Nonminimum Phase Channel

The channel considered has the following impulse response:

$$f(z) = 0.5 + 1z^{-1} - 0.6z^{-2}$$

The linear predictor has the same characteristics as in the previous case.

The nonlinear predictor has one input, 10 neurons in the hidden layer and one output. The parameters are: supervised learning rate equal to 5.10^{-3} and λ equals 10^{-4}. The number of symbols for finding the θ_i was set to 500 symbols and $\kappa = 10^{-7}$. The step size of the AGC equals 10^{-3} and weights in the output layer were initialized at zero and the θ_i were randomly initialized from an uniformly distributed interval: [-1.5,1.5].

Table 1 shows the SER of both predictors, for an SNR of 35 dB. As expected, the linear predictor has a poor performance while the nonlinear one performs pretty well.

	Linear Predictor	Nonlinear Predictor
SER	-3.82 dB	-12.1 dB

Table 1: SER for an SNR of 35 dB
Nonminimum Phase Channel

5. Conclusions

The strategy presented in this paper proposes a nonlinear prediction device based on Artificial Neural Networks. The use of prediction is extended to some cases of nonminimum phase channels and the nonlinear predictor outperforms the linear one even in the cases where it realizes channel equalization.

The division on the learning task in to two steps, a self-organized and a supervised one, avoids reaching points of local minimum in the cost function.

However, this strategy is limited in the situations where the 'valleys' between the Gaussians of the PDF of $x(n-1)$ are deep enough, this deepness is the function of the noise power and channel characteristics. In cases where this condition does not hold, we must consider an adaptation of the previous algorithms, particularly acting on the parameter κ. This improvement is actually under development.

Acknowledgments

The authors would like to thank FUNCAP and CNPq for the partial financial support. They would also like to thank INT-Evry France for the logistical support.

6. References

[1] Haykin, S., *Adaptive Filter Theory.* 3^{rd} edition. Prentice-Hall International, 1996.

[2] Benveniste, A. and M. Goursat, "Blind Equalizers", IEEE Trans. Communication, Vol. COM-32, No. 8, August 1984.

[3] Shalvi O. and E. Weinstein, "New criteria for blind deconvolution of non minimum phase systems", IEEE Trans. on Inform. Theory, Vol. 36, pp. 312-321. 1990.

[4] Macchi, O. "L'égalisation numérique en communications", Annales des Télécomm. 53, No 1-2, pp. 39-58,1998.

[5] Haykin, S., *Neural Networks: A Comprehensive Foundation.* 2^{nd} edition. Prentice Hall International, 1998.

[6] Cavalcante C.C., J.C.M. Mota, B. Dorizzi and J.R. Montalvão, "Preliminary Results in Blind Equalization with Neural Network-Based Prediction", In Proc. XVII Brasilian Telecommunication Symposium, pp.508-513, Vila Velha, Brazil, 1999.