NONLINEAR CHANNEL EQUALIZATION USING MULTILAYER PERCEPTRONS WITH INFORMATION-THEORETIC CRITERION

Deniz Erdogmus[1], Deniz Rende[2], Jose C. Principe[1], Tan F. Wong[2]

[1]Computational NeuroEngineering Lab, [deniz,principe]@cnel.ufl.edu
[2] Wireless Communications Lab, rende@ecel.ufl.edu, twong@ece.ufl.edu
Elect. & Comp. Eng. Dept., University of Florida, Gainesville, FL 32611.

Abstract. The minimum error entropy criterion was recently suggested in adaptive system training as an alternative to the mean-square-error criterion, and it was shown to produce better results in many tasks. In this paper, we apply a multiplayer perceptron scheme trained with this information theoretic criterion to the problem of nonlinear channel equalization. In our simulations, we use a realistic nonlinear channel model, which is encountered when practical power amplifiers are used in the transmitter. The bandwidth-efficient 16-QAM scheme, which uses a dispersed constellation, is assumed.

INTRODUCTION

Since Wiener's classical work on adaptive filters [1], the mean-square-error (MSE) criterion has been the workhorse of function approximation and optimal filtering. It has especially become popular due to the analytical simplicities it introduces when employed to FIR filtering. Recently, we proposed minimizing Renyi's entropy of the error signal in supervised adaptive system training, and used a nonparametric estimator based on Parzen windowing for the entropy [2]. We also know that a system trained with the entropy criterion minimizes an information theoretic distance measure between the probability density functions (pdf) of the desired and the actual outputs [3]. The entropy criterion was applied to a variety of problems including chaotic time series prediction [2,3] and channel equalization [4] with successful results.

The use of large constellations provides bandwidth efficient modulation. Quadrature amplitude modulation (QAM) type modulation techniques have constellations, in which signal points are uniformly spread. Information is carried by both signal amplitude and phase; hence they are not constant envelopes. Thus, efficient nonlinear power amplifiers cannot be utilized in the transmitter, without equalization in the receiver. The use of nonlinear amplifier results in a nonlinear channel. The classical paper by Saleh provides a simple nonlinear model for this nonlinearity introduced by the power amplifiers, which accurately represents experimental data for various practical situations [5]. A variety of approaches employing the MSE criterion have been taken towards solving this nonlinear channel equalization problem. A classical approach suggested by Falconer assumes knowledge of the parametric channel model, and tries to adaptively equalize the nonlinear channel by a suitably chosen equalizer architecture [6]. Decision-feedback is also applied to improve performance. Saleh and Salz, however, proposed a transmitter-based method, where a recursive algorithm is used to predistort the signal constellation to provide a linear overall channel [7]. Recently, the use of neural networks for channel equalization has become popular. An example is [8], where several neural network topologies are compared in terms of both performance and complexity.

The idea of using multilayer perceptrons (MLP) has existed in the literature with successful examples of improved performance over linear equalizers. Motivated by this fact, we employ an MLP based equalization scheme for the nonlinear channel model given by Saleh. In contrast to the above approaches where MSE is adopted as the optimality criterion, the minimum error entropy criterion is utilized in the training process. This choice of optimality criterion is motivated by the improved performance of the neural networks in various applications when compared to the MSE criterion, in the case when the network topology is not sufficient to achieve small error values in training [2,3,4]. In fact, we had proven that if the training error is very small at the optimal solution compared to the kernel size, the error entropy solution is very close to MSE solution. This occurs in two cases: A very large kernel size in entropy estimator, or sufficiently wide span function estimator such that the error is very small [2].

The organization of this paper is as follows. First, we briefly describe the nonlinear channel model and the modulation scheme that is used. Next, we provide an overview of the entropy criterion and the training algorithm for the MLP. Simulation results for the proposed equalizer under additive white Gaussian noise (AWGN) and finally the conclusions are given.

NONLINEAR CHANNEL MODEL

Practical power amplifiers introduce nonlinear distortion in the amplitude and the phase of the transmitted signal. The simple nonlinear model, described by Saleh, is widely used in developing methods to equalize nonlinear channels [5]. This model formulates the amplitude and phase distortion due to a nonlinear amplifier in the transmitter, using two simple two-parameter formulas.

Input signal to the nonlinear channel can be written as

$$s(t) = a(t)\cos[w_c t + j(t)] \tag{1}$$

Here, w_c is the carrier frequency, $a(t)$ is the modulated amplitude, and $j(t)$ is the modulated phase. The amplitude and phase distortion are functions of the amplitude of the input signal, which are denoted by $A[a(t)]$ and $\Phi[a(t)]$ respectively. The output signal after the nonlinear channel is given by

$$r(t) = A[a(t)]\cos\{w_c t + j(t) + \Phi[a(t)]\} \tag{2}$$

The model describes the distortions $A[a(t)]$ and $\Phi[a(t)]$ by the following functions

$$A[x] = \frac{a_a x}{(1 + b_a x^2)} \tag{3}$$

$$\Phi[x] = \frac{a_f x^2}{(1 + b_f x^2)} \tag{4}$$

In this paper, we study the nonlinear channel equalization problem of a communication system employing 16-QAM, which has a rectangular constellation. The transmission signal $s(t)$ for a general M-QAM is given by, in complex baseband representation

$$s_n(t) = \sum_{-\infty}^{\infty} a_n e^{jq_n} p(t - nT) \tag{5}$$

Here n^th^ symbol interval is given by the amplitude and phase a_n and q_n, T is the symbol interval, and $p(t)$ is the pulse waveform with duration T. The data symbol can alternatively be represented by its real and imaginary parts, which can take one of $m = \log_2 M$ values $\pm 1, \pm 3, \cdots, \pm(m-1)$ [9].

The constellation for the 16-QAM is shown below. Bit assignments are chosen as the Gray coding so that neighboring symbols differ only in one bit position. Each symbol corresponds to four data bits.

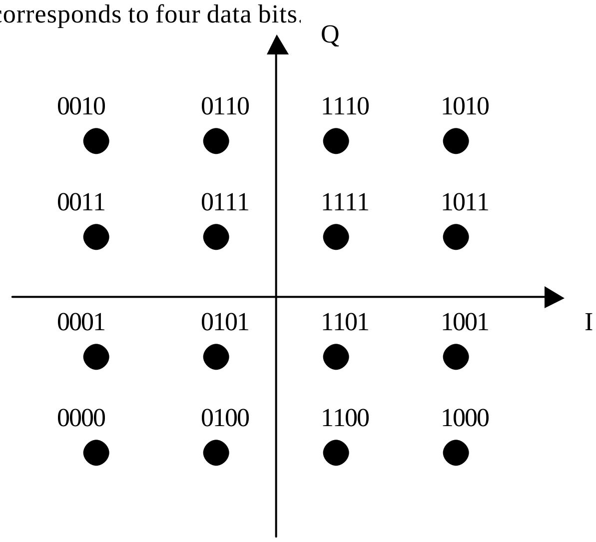


Figure 1. 16-QAM Constellation with Gray coding

The received signal, in complex baseband representation, is composed of the signal distorted by the nonlinear channel and a complex Gaussian noise with uncorrelated real and imaginary parts.

$$r(t) = A[a(t)]e^{j\{j(t)+\Phi[a(t)]\}} + n(t) \tag{6}$$

The goal of the equalizer is to estimate the transmitted symbol from the received signal given in (6).

MINIMUM ERROR ENTROPY ADAPTATION

The general layout of a supervised-learning scheme is illustrated in Fig. 2. Classically, MSE is used as the optimality criterion in the adaptation process. It was shown before that minimizing the error entropy is equivalent to maximizing the mutual information between the desired output and the actual system output [3]. This, in turn, is equivalent to minimizing the α-divergence between the joint probability density functions (pdf) of the input-desired and input-output signal pairs. This quantity is a Riemannian metric, defined by Amari, on the nonlinear manifold of pdfs [10]. This link between error entropy and pdf distances motivates the use of Renyi's entropy as the optimization criterion.

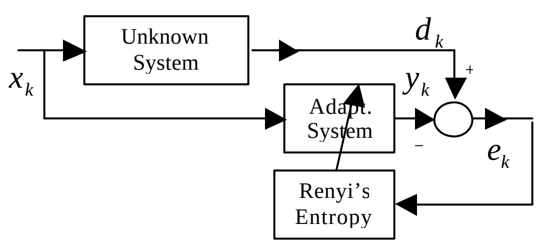


Figure 2. Supervised adaptive system training

In order to proceed with the training process with a finite number of training data, however, we need to devise a nonparametric estimator for entropy. Renyi's entropy for a random variable e is given in terms of its pdf as [11]

$$H_a(e) = \frac{1}{1-a}\log\int_{-\infty}^{\infty} f_e^a(e)de \tag{7}$$

where α >0 is the order of entropy. It is trivial to show, using L'Hopital's rule, that the limit of Renyi's entropy as α→1 is Shannon's entropy. We call the argument of the log the (order-α) information potential [2,12]. It is possible to write the information potential as an expected value as

$$V_a(e) = \int f_e^a(e)de = E\left[f_e^{a-1}(e)\right] \approx \frac{1}{N}\sum_i f_e^{a-1}(e_i) \tag{8}$$

which enables us to obtain the nonparametric estimator we seek after the Parzen window estimator [13] for the pdf of e given below is substituted.

$$V_a(e) = \frac{1}{N^a}\sum_j\left(\sum_i k_s(e_j - e_i)\right)^{a-1} \tag{9}$$

In (9), $k_s(.)$ is the kernel function, usually a symmetric pdf with σ denoting the width. Since log is a monotonic function, minimization of Renyi's entropy corresponds to maximization of the information potential for $\alpha > 1$. Therefore, the information potential can replace the entropy criterion resulting in a simpler cost function.

Suppose the adaptive system under consideration in Fig. 2 is to be trained using a gradient-based algorithm. Then it is required to evaluate the gradient of the information potential estimator in (9) with respect to the weights. The sought expression is

$$\frac{\partial V_a}{\partial w} = \frac{(a-1)}{N^a} \sum_j \left(\sum_i k_s(e_j - e_i) \right)^{a-2} \cdot \left(\sum_i k_s'(e_j - e_i)\left(\frac{\partial y_i}{\partial w} - \frac{\partial y_j}{\partial w}\right)^T \right) \tag{10}$$

The gradient of the output samples with respect to the weights depends on the topology of the adaptive system under consideration. Specifically for a MLP, they can be computed as in the standard backpropagation algorithm [14].

Detailed discussion about the effect of the free parameters α and σ on the structure of the performance surface and the behavior of the algorithm can be found in [2] and [15].

EQUALIZATION METHOD

The proposed equalizer consists of two MLPs operating in parallel. One of them, MLP1, is trained to learn the mapping from the amplitude of the transmitted symbol, $|S|$, to the amplitude of the received signal, $|R|$, where S and R are phasors, obtained from the signals by integrating over one symbol duration and scaling down by the symbol duration. Assigning the input-output variables in this manner also helps the MLP to avoid modeling the noise in the received signal. The other, MLP2, is trained to learn the mapping from $|R|$ to the phase shift introduced by the nonlinear channel, where the desired output is the given by the phase difference $\angle R - \angle S$ between the received and transmitted symbols.

The training process of the MLPs is the depicted in Fig. 3. The training data consists of the amplitudes and angles of the phasors computed as described above.

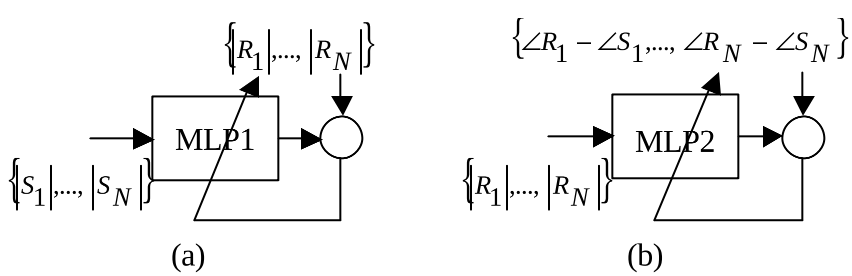


Figure 3. Training of MLPs in the equalizer

In the 16-QAM, there are only three possible values for $|S|$. The outputs of MLP1 corresponding to these three $|S|$ values are compared with the magnitude of the measured phasor, $|R|_m$, and the $|S|$-value that yields the closest estimate is chosen. This is the estimation for the amplitude of the transmitted symbol in the test process. The final decision is made using the estimated $|S|$ and the difference between $\angle R$ and the output of MLP2, which provides an estimate for $\angle S$. The in-phase and quadrature components corresponding to this phasor, formed by the estimated amplitude and phase, are determined. The symbol that minimizes the Euclidean distance to this complex number is decided as the transmitted symbol.

SIMULATION RESULTS

In the previous section, we described the training and testing processes for the equalizer. We trained two MLPs, both with a single hidden layer with 6 neurons and a linear output neuron using the entropy minimization algorithm. The training set consisted of 360 symbols. The variance of the discrete-time noise is adjusted to achieve a predetermined signal-to-noise ratio (SNR) at the equalizer input. SNR here represents the ratio of average bit energy to noise power spectral density (PSD). For each SNR value MLPs are trained and tested independently. In training the MLPs, steepest ascent for information potential was used. A dynamic step size, whose value increases when the update yields a better performance, and decreases when the performance degrades, is utilized. The entropy order was chosen as $a = 3$, and Gaussian kernels with a standard deviation of $s = 1$ were used in Parzen windowing. It was observed that the weights of MLPs converged to the optimal solution in about 20-30 iterations, for all SNR values, with an initial step size of 1. For comparison, we have also trained the same MLPs using the MSE criterion. It was observed that these MLPs converged in 100 iterations starting with the same stepsize.

Upon completion of the training process, the equalizers were tested for bit error rate (BER) using appropriate noise levels and sufficiently long test bit sequences. BER versus SNR plot is shown in Fig. 4. Also shown is the difference in the BER performances of the MLPs trained using MSE and Entropy criteria. As we expected, As the SNR increased, corresponding to smaller training errors, the solutions of the entropy training and MSE training became closer. This fact was proven in [2].

The channel parameters in this simulation were chosen as

$$a_a = 2,\ b_a = 1,\ a_f = p/3,\ b_f = 1. \tag{11}$$

In [7], these values are stated to correspond to a typical severe nonlinear distortion, as we can observe below in Fig. 5. Also in Fig. 5, we illustrate the how the deformation on the signal constellation by the nonlinear channel is reversed by the equalizer. This figure also demonstrates the elimination of the continuous noise in the radial direction by the equalizer. The SNR value is 20dB in this example.

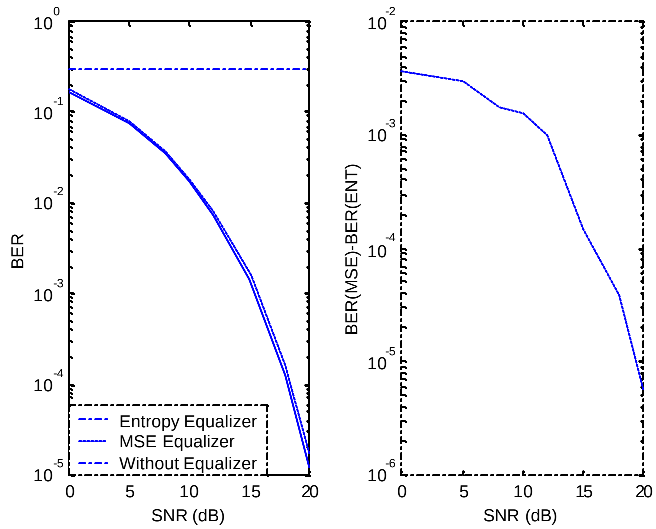
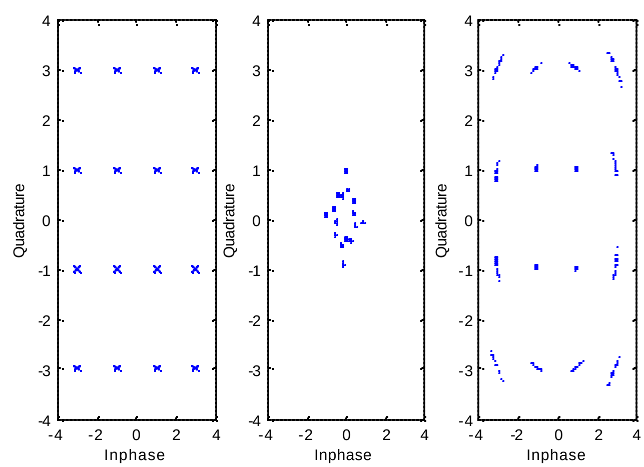


Figure 4. a) BER vs. SNR b) Difference of BER between MSE and Entropy Equalizers



(a) (b) (c)

Figure 5. Constellations for (a) transmitted signals (b) received signals (c) equalizer output

The simulation results presented here clearly show that the MLP-based equalization scheme, which is trained with an information-theoretic criterion, successfully negates the distortion introduced by the nonlinear channel, and furthermore, eliminates the noise component in the radial direction of the constellation. Although we have presented here the simulation results where 360 training symbols were used, our studies with smaller training sets with as few as 60 symbols achieved similar performance in terms of BER.

CONCLUSIONS

Motivated by the improved performance of neural networks trained with an information-theoretic criterion in a variety of problems, and the successful applications of neural networks to the nonlinear channel equalization problem, we have proposed an MLP-based equalization scheme. This scheme is applied to a realistic nonlinear channel model, which delivers a severe distortion on the signal constellation, due to the use of efficient nonlinear power amplifiers in the transmitter. Simulations carried out with the 16-QAM scheme under various SNR conditions pointed out that this deformation in the signal space could be successfully reversed to achieve practically acceptable bit error rates. Also comparison of the entropy equalizer with the MSE equalizer demonstrated the fact that at low SNR values entropy training is more advantageous. Furthermore, it was observed that, as proven in preceding studies, the solutions of the minimum error entropy and minimum mean square error criteria approached to each other as the training error became smaller with increasing SNR.

Some remarkable properties of the proposed equalizer are its computational simplicity, due to the small size of MLPs that can achieve good performance, efficient extraction of information from a small number of training samples, due to the information-theoretic optimality criterion, and the robustness to the radial component of the additive channel noise.

Acknowledgments: This work was supported by the NSF grant ECS-9900394.

REFERENCES

[1] N. Wiener, Extrapolation, Interpolation, and Smoothing of Stationary Time Series with Engineering Applications, Cambridge, MA: MIT Press, 1949.

[2] D. Erdogmus, J.C.Principe, "Generalized Information Potential Criterion for Adaptive System Training," submitted to IEEE Transactions on Neural Networks, Feb. 2001.

[3] D. Erdogmus, J.C.Principe, "An Entropy Minimization Algorithm For Short-Term Prediction of Chaotic Time Series," submitted to IEEE Transactions on Signal Processing, Sept. 2000.

[4] I. Santamaria, D. Erdogmus, J.C. Principe, "Entropy Minimization for Digital Communications Channel Equalization," submitted to IEEE Transactions on Signal Processing, Dec. 2000.

[5] A. Saleh, “Frequency-Independent and Frequency-Dependent Nonlinear Models of TWT Amplifiers,” IEEE Transactions on Communications, vol.29, no.11, Nov. 1981.
[6] D. D. Falconer, “Adaptive Equalization of Channel Nonlinearities in QAM Data Transmission Systems,” Bell System Technical Journal, vo.27, no.7, Sep. 1978.
[7] A. Saleh, J. Salz, “Adaptive Linearization of Power Amplifiers in Digital radio Systems,” Bell System Technical Journal, vo.62, no.4, Apr. 1983.
[8] J.C. Patra, R.N. Pal, R. Baliarsingh, G. Panda, “Nonlinear Channel Equalization for QAM Signal Constellation Using Artificial Neural Networks”, IEEE Transactions on Systems, Man, and Cybernetics, vol.29, no.2, Apr. 1999.
[9] J.G. Proakis, Digital Communications, 3rd ed., NY: McGraw Hill, 1995.
[10] S. Amari, Differential–Geometrical Methods in Statistics, Berlin: Springer-Verlag, 1985.
[11] A. Renyi, Probability Theory, NY: American Elsevier Publishing Company Inc., 1970.
[12] J.C. Principe, D.Xu, J.Fisher, “Information Theoretic Learning,” in Unsupervised Adaptive Filtering, vol I, Simon Haykin Editor, NY: Wiley, 2000, pp 265-319.
[13] E. Parzen, “On Estimation of a Probability Density Function and Mode”, in Time Series Analysis Papers, CA: Holden-Day, Inc., 1967.
[14] D. Rumelhart, G. Hinton, R. Williams, “Learning Internal Representations By Error Backpropagation,” *Nature*, vol. 323, 1986.
[15] D. Erdogmus, J.C. Principe, “Convergence Analysis of the Information Potential Criterion in Adaline Training,” submitted to Neural Networks for Signal Processing 2001, Falmouth, MA.