ADAPTIVE BLIND EQUALIZATION THROUGH QUADRATIC PDF MATCHING

Ignacio Santamaría[1]*, *Carlos Pantaleón*[1], *Luis Vielva*[1], *Jose C. Principe*[2]
[1] DICOM, ETSII y Telecom., University of Cantabria, 39005 Santander, Spain.
Tel: +34 942 201552; fax: +34 942 201488
e-mail:nacho@gtas.dicom.unican.es
[2] CNEL, University of Florida, Gainesville, FL 32611
e-mail:principe@cnel.ufl.edu

ABSTRACT

In this paper we propose a new cost function for blind equalization which aims at forcing a given probability density at the output of the equalizer. In previous works based on this idea, the Kullback-Leibler distance was used as an appropriate measure of the distance between densities. Here we consider the Euclidean (quadratic) distance between the current pdf at the output of the equalizer and the target pdf. Using Parzen windowing with Gaussian kernels for pdf estimation, this quadratic distance can be easily evaluated from data. The adaptive equalization algorithm minimizes the cost function employing a stochastic gradient descent approach.
The algorithm is evaluated in different scenarios through computer simulations, and its performance is compared to that of a minimum Renyi's entropy approach, which is related to the proposed algorithm, and also to the conventional constant modulus algorithm (CMA).

1 INTRODUCTION

The objective of a blind equalizer is to retrieve a digital sequence of symbols sent through an unknown channel, using only the channel output signal and some knowledge of the statistics of the original sequence. In this paper we focus on adaptive (sample-by-sample) blind equalization algorithms due to its simplicity and practical interest. Typically, these algorithms apply stochastic gradient descent (SGD) techniques to some non-MSE cost function which extracts the higher-order statistics of the channel output [1, 2].

To this class of adaptive blind equalization techniques belong the Godard-type algorithms [3], as well as the constant modulus algorithm (CMA) [4], which is a special Godard algorithm and, probably, the most popular blind equalization technique. The main drawback of Godard/CMA equalization algorithms is that they require a long sequence of data to converge. Therefore, some effort to develop new non-MSE cost functions leading to fast and robust adaptive blind equalization algorithms is still needed.

Recently, the authors proposed a new non-MSE cost function based on an efficient nonparametric estimator for Renyi's entropy [5]. In previous works by other authors, the lack of efficient estimators for Shannon's entropy was circumvented by minimizing some measure related to entropy but easier to estimate (such as the normalized kurtosis)[6, 7, 8]. As an alternative to these approaches, the Renyi's entropy estimator described in [9] makes feasible an iterative minimization of the entropy at the output of the equalizer.

However, the cost function proposed in [5] is multimodal and, although for some channels it has shown an increase in convergence speed over the CMA, depending on the initial equalizer settings as well as on the particular channel, it can converge to local minima. Moreover, the number of local minima seems to increase when using multilevel modulations; therefore, its use is restricted to constant modulus signals.

In this paper we propose a new cost function which reduces the number of local minima of [5]. Similarly to [10], the proposed cost function aims at forcing a given probability density at the output of the equalizer. As a measure of distance between densities, the Kullback-Leibler distance is used in [10, 11]. Alternatively, we propose to use a quadratic distance between the current pdf at the output of the equalizer and the desired pdf [9]. Using the Parzen window method for pdf estimation, the quadratic distance can be easily evaluated, yielding a cost function related to that proposed in [5].

Some simulation examples show that, in comparison to [5], the new cost function reduces the number of local minima and also works properly with multilevel modulations. Moreover, in comparison to the CMA, the proposed algorithm still retains the increase in speed of the minimum Renyi's entropy criterion.

2 PREVIOUS WORK

2.1 Problem formulation

We consider a baud-rate sampled baseband representation of the digital communication system. A sequence of i.i.d. complex symbols belonging to a finite alphabet $\{s_k \in R\}$ is sent through a linear time-invariant channel with complex coefficients h_k. The resulting channel output can be expressed as

$$x_k = \sum_n h_n s_{k-n} + e_k,$$

where e_k is a complex zero-mean white Gaussian noise.

The objective of a blind linear equalizer is to remove the intersymbol interference (ISI) at its output without using any training sequence. Typically, the equalizer is designed as an FIR filter with M coefficients $\mathbf{w}$; then, its output is given by

$$y_k = \sum_{n=0}^{M-1} w_n x_{k-n}.$$

* This work has been partially supported by the Spanish Ministry of Science and Technology (MCYT) through grant TIC2001-0751-C04-03.

2.2 Blind equalization with Renyi's entropy

The most popular on-line blind equalization algorithms minimize cost functions of the form

$$J(\mathbf{w}) = E\left[(|y_k|^p - R_p)^2\right], \qquad p = 1, 2, \cdots \tag{1}$$

where $R_p = \frac{E[|s_k|^{2p}]}{E[|s_k|^p]}$ and $E[\cdot]$ denotes mathematical expectation.

In [5] the authors proposed to replace the MSE measure in (1) by an entropy measure. Since an effective nonparametric estimator exist for Renyi's entropy [9], this entropy measure was chosen as an alternative to Shannon's entropy (which, in general, is difficult to estimate), or the use of higher-order moments related to entropy such as the kurtosis [6, 7].

In particular, the proposed cost function is the order-α Renyi's entropy of the modulus signal at the output of the equalizer

$$J_H^p(\mathbf{w}) = H_\alpha\left(|y_k|^p\right), \quad p = 1, 2, \cdots; \tag{2}$$

where the term R_p has been dropped by taking into account that the entropy does not depend on the mean of the signal. In this way, for $p = 1, 2$ the cost function (2) can be considered as an extension of the Sato [12] and CMA [3, 4] cost functions, respectively.

Considering a random variable $Z = |y_k|^p$ with pdf $f_Z(z)$, Renyi's entropy is defined as [9]

$$H_\alpha(Z) = \frac{1}{1-\alpha} \log\left(\int_{-\infty}^{\infty} f_Z(z)^\alpha dz\right). \tag{3}$$

Because the channel is unknown, $f_Z(z)$ cannot be analytically evaluated. However, using a window composed of the current and the past $N-1$ samples: $k+1-N, \cdots, k$; it can be easily estimated by applying the Parzen method with Gaussian kernels. A case of particular interest is that of $\alpha = 2$ (i.e., quadratic entropy); in this situation the cost function reduces to

$$H_\alpha(Z) = -\log\left(\int_{-\infty}^{\infty}\left(\frac{1}{N}\sum_{j=k+1-N}^{k} G_\sigma(z - |y_j|^p)\right)^2 dz\right)$$

where $G_\sigma(y)$ denotes a Gaussian kernel of variance σ^2. Evaluating this integral and taking into account that the logarithm is a monotonic function, blind minimization of Renyi's entropy is achieved by maximizing the following function

$$J_H^p(\mathbf{w}) = \frac{1}{N^2}\sum_{j=k+1-N}^{k}\sum_{i=k+1-N}^{k} G_\sigma(|y_j|^p - |y_i|^p) \tag{4}$$

To avoid convergence to a trivial zero output signal, at each iteration the equalizer coefficients must be constrained somehow; for instance, normalizing the largest tap to one, or fixing the equalizer energy to one.

In [5] it has been shown that, at least for some channels, stochastic gradient maximization of (4) using the smallest possible window ($N = 2$), achieves a remarkable increase in convergence speed in comparison to CMA. However, for other channels and mainly for multilevel modulations, the algorithm tends to get trapped into local minima. In the next section we propose a new cost function which overcomes this drawback.

3 QUADRATIC PDF MATCHING

3.1 Cost function

In order to exploit all the *a priori* statistical information about the problem, the objective of an equalizer should be to force a given probability density at its output. Consider, for instance, the case of a constant modulus signal: in this situation, the equalizer should push the pdf of the random variable $Z = |y_k|^p$ as close as possible to a delta function $f_Z(z) = \delta(z - R_p)$. For any $R_p \neq 0$ the equalizer removes the ISI without gain identification and, after convergence, it is straightforward to estimate the gain.

The idea of equalization via pdf matching is not new. Of particular interest are the criteria proposed in [10], which are based on the Kullback-Leibler (KL) distance between the pdf at the output of the equalizer and the target pdf (see also [11]).

To elaborate on this idea, consider again a constant modulus signal: at the output of the equalizer we have a random variable $Z = |y_k|^p$ with unknown pdf $f_Z(z)$, and a "desired" random variable $Q = R_p + N$, where R_p is a constant and N is a random variable which accounts for the noise at the output of the equalizer. Strictly, N is the modulus of a Gaussian random variable (the noise at the equalizer's input) raised to the pth power. Nevertheless, in order to get a closed-form expression for the cost function, we approximate N as a zero-mean Gaussian with variance σ^2. Although this can be a rather crude approximation for the true pdf of N, it seems to work very well in practice. Therefore, the target pdf is $f_N(q) = G_\sigma(q - R_p)$, and the KL distance between Z and Q is given by

$$D_{KL}(Z||Q) = \int f_Z(z)\ln f_Z(z)dz - \int f_Z(z)\ln G_\sigma(z - R_p)dz. \tag{5}$$

In [10] the first term of (5) (the negative of the Shannon's entropy of Z) is dropped from the cost function, since it can not be easily evaluated or estimated.

It is interesting to point out that by eliminating the first term in (5), and assuming a Gaussian model for the target pdf, the cost function reduces to

$$D_{KL}(Z||Q) = E\left[(|y_k|^p - R_p)^2\right],$$

which, by choosing a constant value $R_p = \frac{E[|s_k|^{2p}]}{E[|s_k|^p]}$, becomes the family of Godard cost functions (1).

It is clear that by dropping the first term in (5) we are also eliminating important statistical information about the pdf of the actual random variable Z. To avoid this drawback we propose to use as a measure of the distance between pdf's, the following quadratic or Euclidean distance, previously described in [9]

$$D_{QD}(Z||Q) = \int \left(f_Z(z) - G_\sigma(z - R_p)\right)^2 dz. \tag{6}$$

Unlike (5), the quadratic distance can be directly estimated using a window of N samples by applying the Parzen windowing method with Gaussian kernels. Specifically, substituting the Parzen window estimate of $f_Z(z)$, developing the square in (6) and integrating, we obtain that the new cost function is given by

$$J^p_{QD}(\mathbf{w}) = \frac{1}{N^2} \sum_{j=k+1-N}^{k} \sum_{i=k+1-N}^{k} G_\sigma(|y_j|^p - |y_i|^p) - \frac{2}{N} \sum_{j=k+1-N}^{k} G_\sigma(|y_j|^p - R_p). \quad (7)$$

This cost function, which has been renamed as $J^p_{QD}(\mathbf{w})$ to make explicit its dependence with the equalizer coefficients and the parameter p, can be related to the entropy cost function as

$$J^p_{QD}(\mathbf{w}) = J^p_H(\mathbf{w}) - \frac{2}{N} \sum_{j=k+1-N}^{k} G_\sigma(|y_j|^p - R_p),$$

note, however, that $J^p_H(\mathbf{w})$ must be maximized, while $J^p_{QD}(\mathbf{w})$ must be minimized.

3.2 Stochastic gradient algorithm

To simplify the derivation of the algorithm we will consider the case in which only the current and the past sample (i.e., $N = 2$) are used to estimate (7). Obviously, this is the more interesting situation from a practical point of view. Additionally, we will derive the algorithm for the particular case $p = 2$. Then, the minimization of (7) using an SGD approach yields the following algorithm

$$\mathbf{w}_{k+1} = \mathbf{w}_k - \mu \left(F_1(y_k, y_{k-1}) + F_2(y_k, y_{k-1}) \right) \quad (8)$$

where μ is the stepsize of the algorithm, and $F_1(y_k, y_{k-1})$ and $F_2(y_k, y_{k-1})$ are given by

$$F_1 = \frac{1}{2} e^{-\frac{(|y_k|^2 - |y_{k-1}|^2)^2}{2\sigma^2}} (|y_k|^2 - |y_{k-1}|^2)(y_{k-1}\mathbf{x}^*_{k-1} - y_k \mathbf{x}^*_k),$$

and

$$F_2 = \sum_{j=k-1}^{k} e^{-\frac{(|y_j|^2 - R_p)^2}{2\sigma^2}} (|y_j|^2 - R_p) y_j \mathbf{x}^*_j.$$

Regarding the selection of R_p, remember that the aim of the proposed cost function is to push the pdf at the output of the equalizer toward *any* constant value, (gain identification can be performed after convergence). Therefore, unlike CMA, in this algorithm R_p does not have to be a constant value fixed in advance. In fact, better performance is obtained if it is adaptively estimated. For constant modulus signal we propose the following updating rule

$$R_p(k) = \lambda R_p(k-1) + (1-\lambda) \frac{|y_k|^2 + |y_{k-1}|^2}{2}, \quad (9)$$

where λ is a value close to one. Similarly, for multilevel signals we propose

$$R_p(k) = \lambda R_p(k-1) + (1-\lambda) \frac{|y_k|^4 + |y_{k-1}|^4}{|y_k|^2 + |y_{k-1}|^2}. \quad (10)$$

At the beginning of the algorithm, the value of R_p is initialized using the channel output according to $R_p = \langle |x_k|^2 \rangle$ or $R_p = \frac{\langle |x_k|^4 \rangle}{\langle |x_k|^2 \rangle}$, respectively; where $\langle \cdot \rangle$ denotes sample mean.

Finally, the proposed algorithm can be summarized in the following steps

1. Initialize μ, λ, σ^2 and $R_p(0)$
2. For $k = 1, 2, \cdots$,
    - 2.1. Update $R_p(k)$ according to (9) or (10).
    - 2.2. Obtain $F_1(y_k, y_{k-1})$ and $F_2(y_k, y_{k-1})$.
    - 2.3. Update the equalizer coefficients using (8).
    - 2.4. Fix the largest tap to one: $\mathbf{w}_k = \mathbf{w}_k / \max(|\mathbf{w}_k|)$

   End.

4 SIMULATION RESULTS

In this section we compare the performance of the proposed quadratic distance cost function J^2_{QD}, the Renyi's entropy cost function J^2_H, and the CMA, in different scenarios. In the first example we assume a QPSK input and consider the following real channel with phase error

$$H_1(z) = e^{j\pi/4}(0.2258 + 0.5161z^{-1} + 0.6452z^{-2} + 0.5161z^{-3}).$$

The channel noise is white and Gaussian for a SNR=30 dB, and a 21-tap equalizer with a tap-centering initialization scheme was applied. As a measure of equalization performance we use the ISI defined by

$$ISI = 10 \log_{10} \frac{\sum_n |\theta_n|^2 - \max_n |\theta_n|^2}{\max_n |\theta_n|^2}$$

where $\theta = \mathbf{h} * \mathbf{w}$ is the combined channel-equalizer impulse response, which is a delta function for a zero-forcing equalizer.

For the entropy and quadratic distance algorithms we used a fixed kernel size $\sigma = 1$, and a fixed value of $\lambda = 0.95$ (used to update the estimate of R_p). The selected stepsizes were $\mu = 0.02$ for J^2_{QD}, $\mu = 0.03$ for J^2_H, and $\mu = 0.006$ for the CMA, which are the largest stepsizes for which the different algorithms converged in all trials. In each case the algorithms were tested in 50 Monte-Carlo trials and the average ISI was plotted in Fig. 1. For this particular channel and in comparison to the CMA, the convergence of the quadratic distance and the entropy algorithms is very fast.

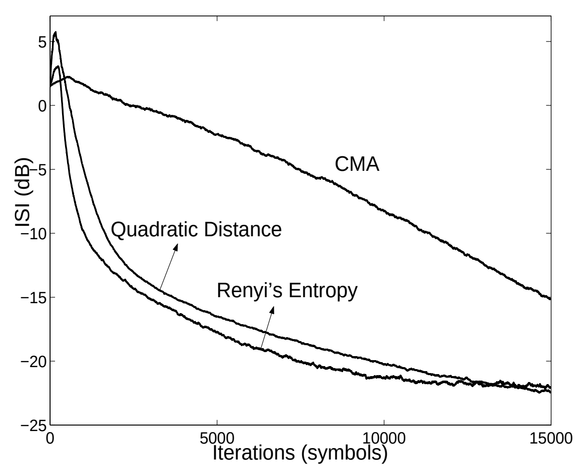

Figure 1: ISI performance of the CMA, the Renyi's entropy ($\alpha = 2$, $p = 2$), and the quadratic distance algorithm ($p = 2$). QPSK input, channel $H_1(z)$ and window size $N = 2$.

In the second example we consider again a QPSK signal and a different channel

$$H_2(z) = \frac{1}{\sqrt{4.75}} \left((0.2 + 0.3j) + (0.9 + 0.9j)z^{-1} + (0.9 - 0.8j)z^{-2} + (0.8j + 0.9)z^{-3} + (0.3 - 0.1j)z^{-4} \right).$$

The SNR is again 30 dB, the equalizer has 17 taps and the stepsize is $\mu = 0.01$ for the three methods. The rest of parameters of the simulation are those of the previous example. The results are depicted in Fig.2: we can see that, in this case, the entropy algorithm converges to a local minimum. On the other hand, the quadratic distance algorithm is still able to converge to the global minimum faster than the CMA. However, the improvement in convergence speed is not as remarkable as in the previous example.

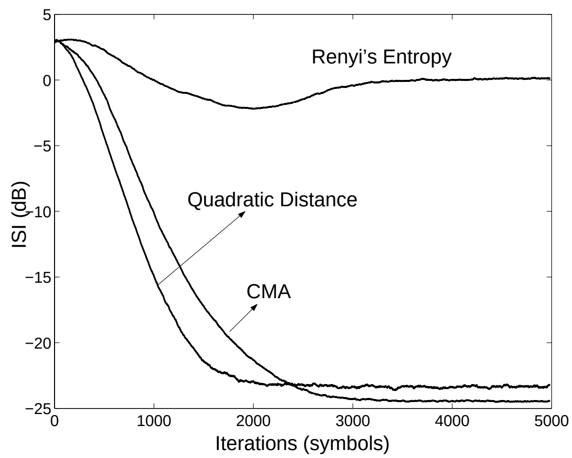

Figure 2: ISI performance of the CMA, the Renyi's entropy ($\alpha = 2$, $p = 2$), and the quadratic distance algorithm ($p = 2$). QPSK input, channel $H_2(z)$ and window size $N = 2$.

In the final example we study the performance of the proposed algorithm with multilevel modulations. We consider a 4-PAM signal with unit power, the channel transfer function is given by $H_3(z) = (0.2258 + 0.5161z^{-1} + 0.6452z^{-2} + 0.5161z^{-3})$, and the final SNR is 30 dB. The selected stepsizes are $\mu = 0.003$ for J^2_{QD}, $\mu = 0.002$ for J^2_H, and $\mu = 0.001$ for the CMA. Fig. 3 shows that the quadratic distance still converges much faster than the CMA, while the entropy criterion converges again to a local minimum.

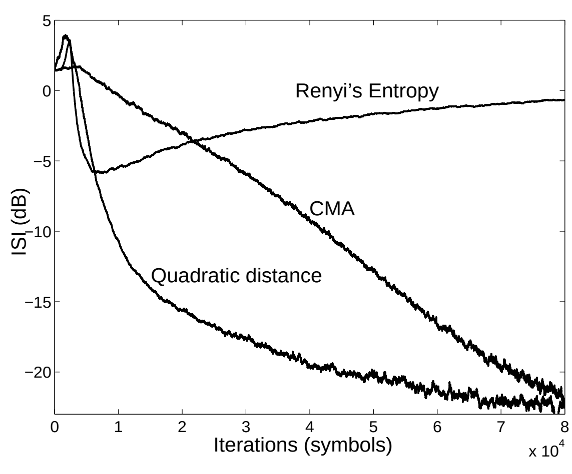

Figure 3: ISI performance of the CMA, the Renyi's entropy ($\alpha = 2$, $p = 2$), and the quadratic distance algorithm ($p = 2$). 4-PAM input, channel $H_3(z)$ and window size $N = 2$.

5 CONCLUSIONS

The quadratic distance between the pdf at the output of the equalizer and a given pdf, based on our statistical knowledge of the input sequence, has been proposed as a new cost function for blind equalization. This quadratic distance is a measure alternative to the typical Kullback-Leibler distance between densities, which has already been applied to blind equalization problems. Unlike the KL distance, the quadratic distance can be directly evaluated from data if the Parzen windowing method with Gaussian kernels is used for pdf estimation. An SGD algorithm is used to minimize the cost function.

It has been shown that the algorithm can be viewed as an extension of the minimum Renyi's entropy algorithm, recently proposed by the authors. In comparison to the latter, the new algorithm seems to have less local minima and, besides, it is not restricted to work with constant modulus signals. In addition, the proposed algorithm is much faster than the CMA.

References

[1] Z. Ding, Y. Li, *Blind Equalization and Identification*, New York: Marcel Dekker, 2001.

[2] S. Haykin Ed., *Blind Deconvolution*, Englewood: Prentice-Hall 1994.

[3] D. N. Godard, "Self-recovering equalization and carrier tracking in two-dimensional data communication systems", *IEEE Trans. on Commun.*, vol. 28, pp.1867-1875, Nov. 1980.

[4] J. R. Treichler, B. G. Agee, "A new approach to multipath correction of constant modulus signals", *IEEE Trans. on Acoust., Speech, Signal Processing*, vol. 31, pp.349-372, Apr. 1983.

[5] I. Santamaria, C. Pantaleon, L. Vielva, J. C. Principe, "A fast algorithm for adaptive blind equalization using order-α Renyi's entropy", submitted to *ICASSP'02*.

[6] D. Donoho, "On minimum entropy deconvolution", *Applied Time Series Analysis II*, pp. 565-609, Academic Press, 1981.

[7] A. T. Walden, "Non-Gaussian reflectivity, entropy and deconvolution", *Geophysics*, vol. 50, pp. 2862-2888, Dec. 1985.

[8] O. Shalvi, E. Weinstein, "Super-exponential methods for blind deconvolution", *IEEE Trans. on Inform. Theory*, vol. 39, pp. 504-519, March 1993.

[9] J. C. Principe, D. Xu, Q. Zhao, J. W. Fisher III, "Learning from examples with information theoretic criteria", *Journal of VLSI Signal Processing*, vol. 26, pp. 61-77, 2000.

[10] J. Sala-Alvarez, G. Vazquez-Grau, "Statistical reference criteria for adaptive signal processing in digital communications", *IEEE Trans. on Signal Processing*, vol. 45, no.1, pp. 14-31, Jan. 1997.

[11] T. Adali, X. Liao, M. K. Sonmez, "Conditional distribution learning with neural networks and its application to channel equalization", *IEEE Trans. on Signal Processing*, vol. 45, no.4, pp. 1051-1064, Apr. 1997.

[12] Y. Sato, "A method of self-recovering equalization for multilevel amplitude modulation", *IEEE Trans. on Commun.*, vol. 23, pp.679-682, Jun. 1975.