A FAST ALGORITHM FOR ADAPTIVE BLIND EQUALIZATION USING ORDER-α RENYI'S ENTROPY

Ignacio Santamaría, C. Pantaleón, L. Vielva *

DICOM, ETSII y Telecom, Univ. of Cantabria
Avda. Los Castros, 39005 Santander, Spain
e-mail: nacho@gtas.dicom.unican.es

Jose C. Principe

Computational NeuroEnginnering Laboratory
Univ. of Florida, Gainesville, FL. 32611
e-mail: principe@cnel.ufl.edu

ABSTRACT

In this paper a novel blind equalization algorithm based on stochastic gradient descent minimization of order-α Renyi's entropy and designed for constant modulus signals is introduced. The algorithm applies a new nonparametric estimator for Renyi's entropy, which has been recently proposed and allows to compute any order of entropy. In comparison with conventional adaptive blind techniques, such us CMA, the proposed algorithm shows a remarkable increase in convergence speed with only a moderate increase in computational cost.

1. INTRODUCTION

Blind adaptive equalizers play a key role in bandlimited digital communications systems in which the transmission of a training sequence is not possible or impractical. For nonminimum-phase channels, it is known that blind techniques must use the higher-order statistics (HOS) of the channel output. They can be explicitly exploited by estimating the higher-order cumulants using a block of received data, or implicitly by minimizing some non-mean square error (MSE) cost function that indirectly extracts the HOS [1, 2]. The latter approach usually leads to simpler algorithms since the minimization of the non-MSE cost function is carried out using stochastic (sample-by-sample) gradient descent (SGD) algorithms.

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. Despite its simplicity, 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.

*This work was partially supported by the European Community and CICYT through the projects TIC2001-0751-C04-03 and 1FD97-1066-C02-01.

Recently, a new information theoretic criterion based on order-α Renyi's entropy has been proposed and successfully applied to a number of applications including blind source separation [5] and data-aided equalization [6]. In these works the interest was placed on quadratic entropy ($\alpha = 2$) estimated over a block of data (batch) and using Parzen windowing with Gaussian kernels. However, the technique has been generalized to cope with any order of entropy, and the minimization can be carried out on a sample-by-sample basis [7].

Although the use of some measure related to entropy (such as the normalized kurtosis) has been widely applied to blind equalization and deconvolution [8, 9, 10], we are not aware of any equalization technique using directly the entropy as a cost function. This is due to the fact that the most widely known definition of entropy (i.e., Shannon's entropy) is, in general, hard to estimate and minimize. As it has been shown, the use of Renyi's entropy overcomes this problem.

In this paper we propose and study the application to blind equalization of this new family of algorithms based on SGD minimization of order-α Renyi's entropy. In comparison with conventional adaptive blind techniques, such us CMA, the proposed algorithms show a remarkable increase in convergence speed with only a moderate increase in computational cost.

2. A NEW FAMILY OF COST FUNCTIONS

In this paper we focus on blind equalization of constant modulus signals; in particular, in this section we will assume a QPSK signal. The channel output can be described using the following discrete-time baseband representation

$$x_k = \sum_{n=0}^{l_c} h_n s_{k-n} + e_k, \qquad (1)$$

where s_k is assumed to be a sequence of i.i.d. complex symbols, h_k are the complex channel coefficients (we assume here an FIR channel), and e_k is a 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

$$y_k = \sum_{n=0}^{M-1} w_n x_{k-n}. \quad (2)$$

The most popular blind algorithms are the family of Godard algorithms [3], which are stochastic gradient descent methods for minimizing the cost functions

$$J_G(\mathbf{w}) = E\left[(|y_k|^p - R_p)^2\right], \quad p = 1, 2, \cdots \quad (3)$$

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

For the particular case $p = 2$, Eq.(3) is the cost function of the CMA, which was independently developed using the idea of penalizing the output samples that do not have the constant modulus property [4]. Using an SGD minimization approach, the CMA can be written as

$$\mathbf{w}_{k+1} = \mathbf{w}_k - \mu(|y_k|^2 - R_2) y_k \mathbf{x}_k^* \quad (4)$$

where the superscript * denotes complex conjugate.

Due to its simplicity the CMA is one of the most widely used blind equalization techniques, although typically it requires a large number of samples to converge satisfactorily.

In this paper we propose the following generalization of the CMA cost function: instead of penalizing the squared error deviations from the desired constant modulus property, we penalize the entropy of the deviations. Specifically, we use order-α Renyi's entropy, which, for a random variable e with pdf $f(e)$, is defined as [11]

$$H_\alpha(e) = \frac{1}{1-\alpha} \log\left(\int_{-\infty}^{\infty} f(e)^\alpha de\right). \quad (5)$$

The entropy of a signal is a function of its pdf. In this way, by minimizing it we are using much more information than by minimizing just its variance. Similarly to other nonlinear problems where this criterion has been applied, it is expected that the extra information exploited by the entropy will provide some advantage, mainly in terms of speed of convergence.

In summary, we propose to use the following family of entropy-based cost functions

$$J_\alpha^p(\mathbf{w}) = H_\alpha(|y_k|^p - R_p), \quad p = 1, 2, \cdots \quad (6)$$

which, since the entropy does not depend on the mean of the signal, is equivalent to

$$J_\alpha^p(\mathbf{w}) = H_\alpha(|y_k|^p), \quad p = 1, 2, \cdots \quad (7)$$

Therefore, the new family of cost functions is the order-α Renyi's entropy of the equalizer's output raised to the pth power. In the sequel we will concentrate on the case $p = 2$, which can be considered as an extension of the CMA cost function.

For a doubly infinity equalizer and considering a constant modulus input and a noiseless situation, the minima of (7) are obtained when $f(|y_k|^p) = \delta(|y_k|^p - K)$ for any K, i.e., when the equalizer output is also a constant modulus signal. Except for the case $K = 0$, which can be easily avoided by constraining the equalizer parameters (forcing the central tap to one, for instance), the pdf of the input sequence and that of the equalizer's output coincide up to a gain factor. Then, the minima of (7) correspond to points where perfect equalization (zero-forcing) is attained.

3. MINIMIZATION OF THE COST FUNCTION

The minimization of the entropy cost function (7) is equivalent, for $\alpha > 1$, to maximize the function

$$V_\alpha(\mathbf{w}) = E\left[f(|y_k|^2)^{\alpha-1}\right], \quad (8)$$

which is called the information potential [11]. Using a window of N samples formed by the current and the past $N-1$ outputs of the equalizer, the information potential can be estimated by substituting the expectation by a sample mean [7]

$$V_\alpha(\mathbf{w}) \approx \frac{1}{N} \sum_j f(|y_j|^2)^{\alpha-1}. \quad (9)$$

If the pdf of $|y_j|^2$ is estimated using the Parzen window method, we finally get that the objective function to be maximized in our problem is

$$V_\alpha(\mathbf{w}) = \frac{1}{N^\alpha} \sum_j \left(\sum_i G_\sigma(|y_j|^2 - |y_i|^2)\right)^{\alpha-1}, \quad (10)$$

where $G_\sigma(y)$ denotes a Gaussian kernel of variance σ^2.

To maximize (10), the coefficients of the equalizer are updated as

$$\mathbf{w}_{k+1} = \mathbf{w}_k + \mu \frac{\partial V_\alpha(\mathbf{w})}{\partial \mathbf{w}} \quad (11)$$

where μ is the stepsize of the algorithm. Finally, the derivative of the information potential with respect to the equalizer coefficients is

$$\frac{\partial V_\alpha}{\partial \mathbf{w}} = \frac{\alpha-1}{\sigma^2 N^2} \sum_{j=k+1-N}^{k} f(|y_j|^2)^{\alpha-2} F(y_j) \quad (12)$$

where the term $F(y_j)$ is given by

$$F(y_j) = \sum_{i \neq j} G_\sigma(|y_j|^2 - |y_i|^2)(|y_j|^2 - |y_i|^2)(y_i \mathbf{x}_i^* - y_j \mathbf{x}_j^*). \quad (13)$$

As we can see from (13), each data sample y_j interacts with all other samples in the block of data used in the current iteration, pushing the solution towards a constant modulus signal. Moreover, in (12) the individual forces $F(y_j)$ are weighted by the probability of that sample (measured through its pdf estimate) raised to the power $\alpha - 2$.

For $\alpha = 2$ (quadratic entropy), Eq. (12) reduces to

$$\frac{\partial V_\alpha}{\partial \mathrm{w}} = \frac{1}{\sigma^2 N^2} \sum_j F(y_j), \tag{14}$$

from a practical standpoint, a case of particular interest consists of using just the current and the past sample in the estimation of $F(y_j)$. In this situation ($\alpha = 2$, $N = 2$), the quadratic entropy algorithm is given by

$$\mathrm{w}_{k+1} = \mathrm{w}_k + \frac{\mu}{2\sigma^2} G_\sigma(\ _k)\ \ _k (y_{k-1}\mathrm{x}^*_{k-1} - y_k \mathrm{x}^*_k), \tag{15}$$

where $\ _k = |y_k|^2 - |y_{k-1}|^2$. The computational cost of this adaptive algorithm is similar to that of the CMA.

The extension of this algorithm to a general case $\alpha \neq 2$ requires estimating the pdf of $|y_j|^2$. Although this estimation can be carried out using the Parzen window method, a simpler alternative consists of assuming a Gaussian model

$$f(|y_j|^2) = \frac{1}{\sqrt{2\pi}\sigma_y} \exp - \left(\frac{(|y_j|^2 - \mu_y)^2}{2\sigma_y^2} \right), \tag{16}$$

in this way, only the mean μ_y and the variance σ_y^2 of the squared output of the equalizer must be estimated. Moreover, these estimates can be easily updated on a sample-by-sample basis.

4. RESULTS

In the first example we assume a QPSK input and consider the following real channel with phase error

$$H_1(z) = \frac{e^{j\pi/4}}{1.41}(04 - 0.6z^{-1} + 1.1z^{-2} - 0.5z^{-3} + 0.1z^{-4}).$$

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

$$ISI = 10 \log_{10} \frac{\sum_n |\theta_n|^2 - \max_n |\theta_n|^2}{\max_n |\theta_n|^2} \tag{17}$$

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

First, we have tested the proposed adaptive blind equalization algorithm for quadratic entropy ($\alpha = 2$) and different data-block sizes ($N = 2, 5, 10$). We used a fixed stepsize $\mu = 0.03$ and a fixed kernel size $\sigma = 1$ for entropy estimation. In each case the algorithm was tested in 25 Monte-Carlo trials and the average ISI was plotted in Fig. 1. For comparison purposes, the corresponding performance of the CMA is included; in this case, a value of $\mu = 0.009$ was used, which is the largest stepsize for which the CMA converged in all trials.

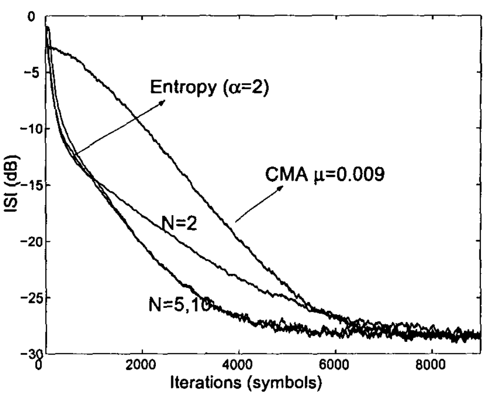

Fig. 1. ISI performance of the CMA and the proposed algorithm using $\alpha = 2$ and different data-block sizes. QPSK input and channel $H_1(z)$.

In comparison to the CMA, the increase in convergence speed is remarkable. More important is that the algorithm has a very fast initial convergence and slows down later; therefore, the switch to a simpler and faster decision-directed mode can be made much earlier than with the CMA. For instance, Fig. 2 shows the complex channel output and the equalizer output using the proposed algorithm with $\alpha = 2$ and $N = 2$ after 3000 symbols. It is also remarkable that most of the improvement in speed is obtained using the smallest window size, $N = 2$, thus requiring only a small increase in computational cost with respect to the CMA. The learning curves for $N = 5$ and $N = 10$ are very similar.

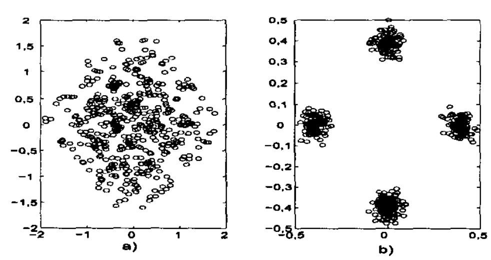

Fig. 2. a) Complex channel output, and b) proposed equalizer output with $\alpha = 2, N = 2$ after 3000 symbols.

We have studied the effect of α in the algorithm: Fig. 3 shows the average ISI for the proposed algorithm using $N = 2$ and three different values for the entropy order $\alpha = 2$, 2.4 and 2.8. A value $\alpha > 2$ can achieve a small increase in speed, but this increase comes when the eye-pattern is already sufficiently open to a degree where a decision-directed mode can operate satisfactorily. Moreover, this increase in speed does not compensate the increase in computational cost, therefore a value $\alpha = 2$ is recommended. In the second example the input sequence is a 2-PAM sig-

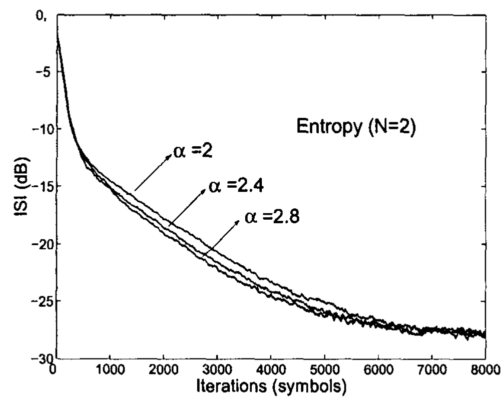

Fig. 3. ISI performance of the proposed algorithm using $N = 2$ and three different values for the entropy order.

nal and the channel is: $H_2(z) = (02258 + 0516z^{-1} + 0.6452z^{-2} - 0.516z^{-3})$. The SNR is 30 dB and the equalizer has 19 coefficients. Again, a fixed stepsize $\mu = 0.03$ and a fixed kernel size $\sigma = 1$ was used for the entropy algorithm, while a value of $\mu = 0.007$ was selected for the CMA. Fig. 3 shows the average learning curve for this example: the improvement in convergence speed is clear. Moreover, for $N = 5$ and $N = 10$ the final result is even better than that of the CMA.

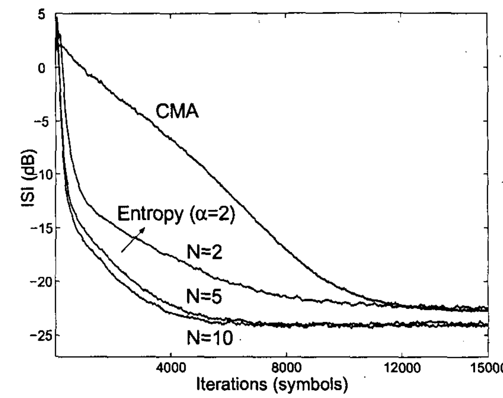

Fig. 4. ISI performance of the CMA and the proposed algorithm using $\alpha = 2$ and different data-block sizes. 2-PAM input and channel $H_2(z)$.

5. CONCLUSIONS

A new family of cost functions based on order-α Renyi's entropy has been applied to blind equalization of constant modulus signals. They can be minimized using SGD techniques, thus leading to simple and useful algorithms. In particular, the use of quadratic entropy ($\alpha = 2$) and the shortest window ($N = 2$) is recommended, since in this case an algorithm with a computational cost similar to that of the CMA, but with a much faster convergence, is obtained. Although it has been shown that for a doubly infinity equalizer the minima of the proposed cost function correspond to points where perfect equalization is attained, a rigorous convergence analysis is still needed. Furthermore, the extension of these ideas to multilevel modulations (QAM, in particular) is an interesting future research line.

6. 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] K. E. Hild, D. Erdogmus, J. C. Principe, "Blind source separation using Renyi's mutual information", *IEEE Signal Processing Letters*, vol. 8, pp. 174-176, June 2001.

[6] I. Santamaria, D. Erdogmus, J. C. Principe, "Entropy minimization for digital communications channel equalization", submitted to *IEEE Trans. on Signal Processing*, 2000.

[7] D. Erdogmus, J. C. Principe, "Convergence analysis of the information potential criterion in adaline training", *Proc. of NNSP*, pp. 123-132, Falmouth, MA, Sept. 2001.

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

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

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

[11] 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.