BLIND EQUALIZATION BY SAMPLED PDF FITTING

M. Lázaro, I. Santamaría, C. Pantaleón[*]

DICOM, ETSIIT, Univ. of Cantabria
Av Los Castros s/n, 39005 Santander, Spain
{marce,nacho,carlos}@gtas.dicom.unican.es

D. Erdogmus, K. E. Hild II, J. C. Principe[†]

Computational Neuroengineering Lab.
Univ. of Florida, Gainesville, FL 32611
{deniz,hildk,principe}@cnel.ufl.edu

ABSTRACT

This paper presents a new blind equalization technique for multilevel modulations. The proposed approach consists of fitting the probability density function (pdf) of the corresponding modulation at a set of specific points. The symbols of the modulation, along with the requirement of unity gain, determine these sampling points. The underlying pdf at the equalizer output is estimated by means of the Parzen window method. A soft transition between blind and decision directed equalization is possible by using an adaptive strategy for the kernel size of the Parzen window method. The method can be implemented using a stochastic gradient descent approach, which facilitates an on-line implementation. The proposed method has been compared with CMA and Benveniste-Goursat methods in QAM modulations.

1. INTRODUCTION

Channel equalization is one of the most common and successful applications of adaptive filters in digital communication systems. When the transmitter sends a training sequence, the filter coefficients can be easily adapted using the LMS algorithm. However, in many digital communication systems, the transmission of a training sequence is not possible or it is very costly. In these cases, blind equalization algorithms, which do not need a reference sequence, are necessary.

There exist a number of different blind algorithms [1, 2]. In communication systems, the Constant Modulus Algorithm (CMA) [3] is the most commonly employed algorithm. It belongs to the family of Godard algorithms [4]. One of its more remarkable features is its simplicity. Its main drawback is that it usually requires a high number of data symbols to achieve convergence.

Several attempts have been made to improve the convergence speed of conventional blind techniques. For instance, Renyi's entropy has been introduced as a cost function for blind equalization [5]. This approach uses an efficient nonparametric estimator of this entropy measurement based on Parzen window method to estimate the underlying pdf [6]. Although this method provides excellent results for some channels, it fails to equalize other ones. In addition, it does not work for multilevel modulations.

An interesting alternative consists in trying to force the probability density at the output of the equalizer to match the known constellation pdf. The well known Kullback-Leibler divergence between densities has been employed as a cost function for linear [7] and nonlinear (neural networks based) equalizers [8]. This method has been tested in binary PAM modulations. A new method based on the quadratic distance between pdf's has been proposed in [9]. It has the advantage of being much simpler than [7], facilitating the on-line implementation. However, it was designed only for constant modulus algorithms (a simple Gaussian model is used for the target pdf to be fitted).

In this paper, we propose a new blind algorithm that aims at forcing the probability density of the equalizer output at several sampling points. It admits a stochastic gradient-based implementation, which is similar in complexity to CMA. Therefore, it can be easily implemented in on-line applications. Moreover, it allows a soft switch between blind and decision directed equalization, unlike CMA-based systems. The Benveniste-Goursat algorithm [10], which is used for instance in digital TV systems [11], implements a similar soft transition. The proposed method has been successfully tested in quadrature amplitude modulations (QAM).

2. BLIND EQUALIZATION FORMULATION

In a communication system framework, the blind equalization problem can be formulated as follows: a sequence $\{s_k\}$ of i.i.d. complex symbols belonging to the constellation of any digital modulation is sent through a channel. Usually the channel is described by means of its discrete time complex coefficients h_k (assuming a FIR channel). Therefore,

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

[†]This work is partially supported by NSF grant ECS-9900394

the channel output is obtained by

$$x_k = \sum_{n=0}^{L_h-1} h_n s_{k-n} + e_k, \tag{1}$$

where the noise sequence e_k is typically modeled by a zero-mean white gaussian noise process.

In this approach we have only considered linear equalization. A linear equalizer, implemented by means of a FIR filter, will be used to minimize the intersymbol interference introduced by the channel. Hence, the equalizer output is

$$y_k = \sum_{n=0}^{L_w-1} w_n x_{k-n} = \mathbf{w}^{\mathrm{T}} \mathbf{x_k}, \tag{2}$$

where $\mathbf{w}$ is the vector of filter coefficients. The goal of the blind algorithm is to find the coefficients that minimize the intersymbol interference (ISI) introduced by the channel. Since it does not have a reference sequence, a blind algorithm must to make use of some *a priori* knowledge of the statistics of the input signal to adapt the weights. For instance, the Godard algorithms [4] minimize the following cost function

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

where the *a priori* knowledge is carried in the ratio

$$R_p = \frac{E[|s_k|^{2p}]}{E[|s_k|^p]}. \tag{4}$$

CMA is the specific Godard algorithm for $p = 2$.

3. SAMPLED PDF COST FUNCTION

In order to employ as much information as possible, in this paper we propose to make use of the *a priori* knowledge of the probability density function (pdf) of $S^p = \{|s_k|^p\}$. This pdf contains more information than the ratio (4). Specifically, the aim is to fit the pdf at a number, N_p, of representative points (the sampling points). In this way, the proposed cost function is

$$J(\mathbf{w}) = \frac{1}{N_p} \sum_{i=1}^{N_p} \left(f_{Y^p}(r_i) - T_i\right)^2, \tag{5}$$

where $Y^p = \{|y_k|^p\}$, $f_X(x)$ denotes the pdf of X at x, and T_i are the target values of the pdf at r_i ($T_i = f_{S^p}(r_i)$). The minimum of $J(\mathbf{w})$ is obtained when the pdf of Y^p matches $f_{S^p}(r)$ at the sampling points r_i.

The pdf at the equalizer output is estimated by using the Parzen window method. Considering a window of the L previous symbols, the estimate of f_{Y^p} at radius r_i at time k is

$$\hat{f}_{Y^p}(r_i^p) = \frac{1}{L} \sum_{j=0}^{L-1} K_\sigma(r_i - |y_{k-j}|^p), \tag{6}$$

where $K_\sigma(x)$ is the Parzen window kernel of size σ. Gaussian kernels with standard deviation σ are employed. For the sake of consistency, the target pdf values (the values at r_i) must be evaluated taking into account the nature of the estimator we are using to estimate $f_{Y^p}(r_i)$. Consequently, the original pdf must be convolved with the same kernel used by the estimator

$$T_i = \frac{1}{N_s} \sum_{j=0}^{N_s-1} K_\sigma(r_i - |s_j|^p), \tag{7}$$

where N_s is the number of complex symbols in the constellation of the specific modulation.

Substituting (7) and (6) into (5), the cost function at instant k becomes

$$J(\mathbf{w}_k) = \frac{1}{N_p} \sum_{i=1}^{N_p} \left(\frac{1}{L} \sum_{j=0}^{L-1} K_\sigma(r_i - |y_{k-j}|^p) - T_i \right)^2. \tag{8}$$

4. STOCHASTIC GRADIENT ALGORITHMS

The gradient of $J(\mathbf{w}_k)$ with respect to the weight vector is

$$\frac{dJ(\mathbf{w}_k)}{d\mathbf{w}_k} = -\frac{2}{N_p} \sum_{i=1}^{N_p} \left(\frac{1}{L} \sum_{j=0}^{L-1} K_\sigma(r_i - |y_{k-j}|^p) - T_i \right) \left(\frac{1}{L} \sum_{j=0}^{L-1} K'_\sigma(r_i - |y_{k-j}|^p) \frac{d(|y_{k-j}|^p)}{d\mathbf{w}_k} \right). \tag{9}$$

This expression can be used to implement a batch version of the algorithm. However, for the sake of minimizing the computational burden in order to allow the online implementation of the method, a stochastic gradient algorithm has been developed considering only the actual sample ($L = 1$). We have also considered $p = 2$, which is the more convenient choice. In this case, the estimation of the pdf is performed using only one term of the Parzen sum (6). Therefore, the same kind of estimation must be employed for the target T_i in order to guarantee consistency. On the other hand, the natural choice for the sampling points r_i is $|s_i|^2$ (this point is discussed in section 5.1). In this case, the more suitable target value is given by

$$T_i = K_\sigma(r_i - |s_i|^2) = K_\sigma(0), \tag{10}$$

and the adaptation term of the stochastic algorithm becomes

$$\Delta\mathbf{w}_k = -\frac{2}{N_p} \sum_{i=1}^{N_p} \left(K_\sigma(r_i - |y_k|^2) - K_\sigma(0)\right) K'_\sigma(r_i - |y_k|^2) y_k \mathbf{x}_k^*. \tag{11}$$

Finally, the weights are adapted by

$$\mathbf{w}_{k+1} = \mathbf{w}_k - \mu_\sigma \Delta \mathbf{w}_k. \tag{12}$$

The normalized step size $\mu_\sigma = \mu\sigma^3$ has been introduced to compensate the $1/\sigma^3$ term of the gaussian kernel derivative $K'_\sigma(x)$. Therefore, in the following we will only consider μ when referring to the step size.

5. IMPLEMENTATION DETAILS

5.1. Selection of the sampling points r_i

As we have already said, the natural choice for r_i is $|s_i|^2$, taking only the values that are different. For instance, in a 16 QAM (with $\{\pm 1, \pm 3\}$ in both the real and the imaginary parts of each symbol), the number of sampling points is $N_p = 3$, with $|s_i|^2 = \{2, 10, 18\}$.

This choice provides equalization with exact gain identification when using a small kernel size. In this case, a sample only interacts with the closest target and it is easy to understand that the minimum of $J(\mathbf{w})$ is obtained when the equalizer pdf fits the constellation pdf. However, a large kernel size allows the interaction of each sample with all the target values. In this case, the minimum of the cost function (for $L = 1$) is achieved when the pdf at the equalizer output is slightly scaled down. This means that this choice for the sampling points equalizes the channel up to gain identification. In order to ensure unity gain, the adapting expression (11) must be analyzed. Naturally, we have to require that the expectation of (11) be zero when perfect equalization is achieved. To obtain an analytical expression for r_i to accomplish this requirement is rather involved. However, it is simple to obtain a numerical solution. Under the assumption of perfect equalization and unity gain, it is possible to show that the sampling points for which the expectation of (11) equals to zero are

$$r_i = F(\sigma)|s_i|^2, \quad i = 1, \cdots, N_p, \tag{13}$$

where $F(\sigma)$ is a compensation factor, which depends on the kernel size. For a 16QAM, this factor has been obtained numerically and it is plotted in Figure 1.

It can be seen that the compensation factor becomes 1 for small kernel sizes. This means that in this situation, when only iteration with the closer target is allowed, the sampling points are $|s_i|^2$, which is the more intuitive choice.

We want to note that, although strictly speaking the target values given by (10) should be modified to take into account the deviation with respect to $|s_i|^2$, in practice we have observed that using $K_\sigma(0)$ provides more satisfactory results.

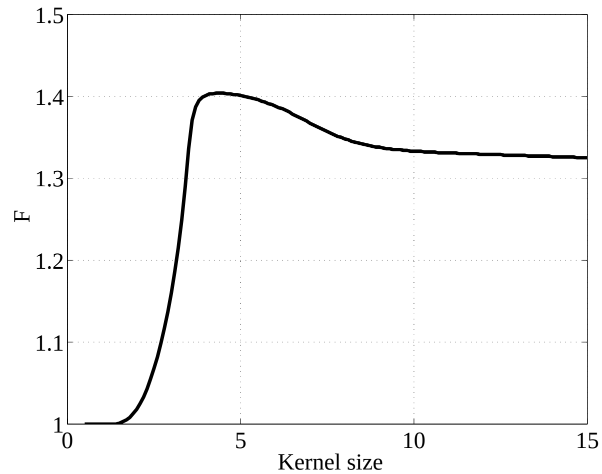

Fig. 1. Compensation factor for a 16 QAM

5.2. Effects of the kernel size

The kernel size σ plays a key role in this algorithm. It determines both the convergence speed and the accuracy of the final solution. A large kernel size provides a fast convergence because it allows the interation of the samples with all the target values. On the other hand, a small kernel size, which only allows iteration with the closest target, is more appropriate when the concern is accuracy. This limited interaction produces, in practice, a decision directed equalization behavior.

In communication systems, convergence speed is a very important issue. Typically, blind algorithms reduce ISI until the eye of the constellation is opened. At this point, the system switches to decision directed equalization, which produces a finer equalization. To obtain a fast convergence a large kernel size must be selected. When convergence is achieved, it is possible to switch to decision directed equalization or to use a small kernel size. However, the proposed method allows another interesting alternative: a soft transition from blind to decision directed equalization by adaptively controlling the kernel size.

5.3. Soft blind to decision directed transition

We have discussed the fact that the kernel size controls the convergence speed and the final accuracy of the solution, and the fact that the two are in opposition with each other. Therefore, an interesting approach consists of adaptively controling the kernel size to have a large value in the blind stage and to progressively decrease it during convergence to obtain a finer final equalization.

An error measure, which is based on the variance of the error with respect to the closer target, has been employed to control the kernel size. This measure is iteratively adapted, using a forgetting factor α, by means of

$$E_{k+1} = \alpha E_k + (1-\alpha) \min_{\{i=1,\cdots,N_p\}} \left((|y_k|^2 - r_i^2)^2\right). \quad (14)$$

In this approach, the kernel size is obtained by

$$\sigma_k = aE_k + b, \quad (15)$$

where a and b are empirically determined constants. For instance, for a 16 QAM modulation we have found, after testing in a large number of channels, that $a = 3.5$ and $b = -9.5$ provide very good results.

We would want to remark that in order to obtain a suitable soft transition, the sampling points r_i must be adapted at each iteration according to the current kernel size. In this approach, a look-up table has been used to evaluate $F(\sigma)$.

Taking this into account, the soft transition algorithm can be summarized in the following steps:

1. Initialize μ, E_1, α.
2. For k=1,2,...
   (a) Evaluate σ_k by (15)
   (b) Update $\mu_\sigma = \mu\sigma_k^3$
   (c) Obtain $F(\sigma)$
   (d) Update $r_i = F(\sigma)|s_i|^2$
   (e) Evaluate $\Delta\mathbf{w}_k$ by (11)
   (f) Update $\mathbf{w}_{k+1}$ by (12)
   (g) Estimate E_{k+1} by (14)

   End

6. RESULTS

The proposed method has been tested in several different channels for QAM modulations. The intersymbol interference (ISI) will be used as a figure of merit to compare the performance of the methods. It can be computed as

$$ISI = 10\log_{10}\frac{\sum_n |\theta_n|^2 - \max_n |\theta_n|^2}{\max_n |\theta_n|^2}, \quad (16)$$

where $\theta = \mathbf{h} * \mathbf{w}$ is the combined channel-equalizer impulse response.

6.1. Blind equalization

In this section we will compare the performance of the proposed algorithm working only in a blind mode. Results will be compared with those provided by CMA, which is the most widely used blind algorithm. The main requirement of blind equalization is convergence speed. This means that a large kernel size is necessary.

In the first example we have considered a 16 QAM modulation ($\pm\{1,3\}$ levels for in phase and quadrature components) and the following channel

$$H_1(z) = (0.2258 + 0.5161z^{-1} + 0.6452z^{-2} + 0.5161z^{-3}).$$

White Gaussian noise, with a signal to noise ratio (SNR) of 30 dB, has been added at the output of the channel. A $L_w = 21$ taps equalizer with tap-centering initialization has been employed. The following parameters have been selected: a step size $\mu = 0.02$ and a kernel size $\sigma = 15$. This large kernel size has been selected in order to maximize the convergence speed. Figure 2 compares the average results obtained in 100 Montecarlo trials with those obtained with the CMA algorithm using the step size μ =1e-5 (the value that provides the fastest stable convergence).

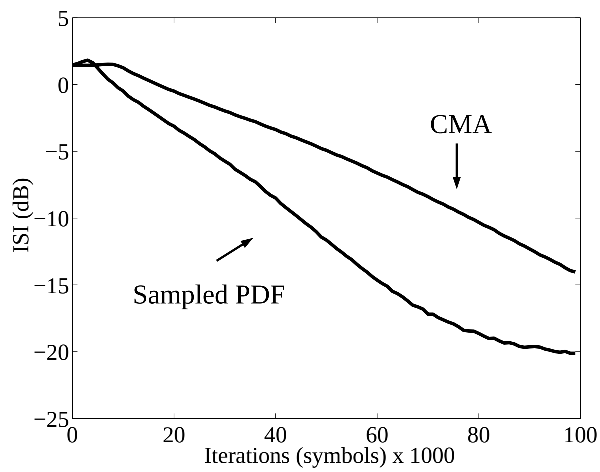

Fig. 2. Convergence for a 16 QAM in $H_1(z)$

It can be seen that the sampled-pdf approach exhibits a remarkably faster convergence than CMA. In this channel, CMA shows a very slow convergence, but the proposed approach has shown to be faster in channels where CMA provides a fast convergence, even when a higher level modulation is used. For instance, we have also tested the following non-minimum phase channel

$$H_2(z) = \frac{1}{\sqrt{4.75}}\left[(.2+.3i) + (.9+.9i)z^{-1} + (.9-8i)z^{-2} + (.8+.9i)z^{-3} + (.3-.1i)z^{-4}\right]. \quad (17)$$

In this case a 256 QAM modulation ($\pm\{1,3,...,15\}$ levels for in phase and quadrature components) has been tested . The kernel size is $\sigma = 450$ (note that we have to ensure the interaction of the symbols with all the target values and the largest value for $|s_i|^2$ is 450). The step sizes $\mu = 1e-7$ and

$\mu = 1e-3$ have been selected for CMA and sampled-pdf respectively, which are the values that provide the fastest stable convergence in both cases. The convergence results are plotted in Figure 3.

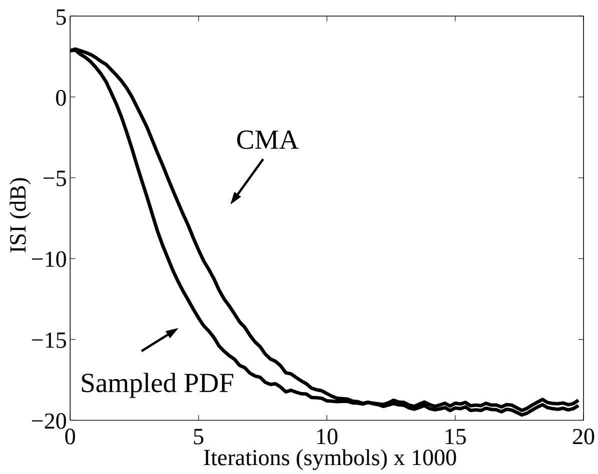

Fig. 3. Convergence for a 256 QAM in $H_2(z)$

Again, the sampled-pdf shows faster convergence. We want to remark that both methods have been tested over a large number of channels and not a single case has been found where CMA outperforms the proposed approach in terms of convergence speed.

6.2. Decision directed equalization

One of the features of the proposed approach is that it allows a fine equalization by selecting a small kernel size. For instance, Figure 4 shows the ISI evolution using $\sigma = 1$ for $H_2(z)$, after the initial convergence has been achieved using $\sigma = 15$ for a 16 QAM modulation. The proposed method has been compared with decision directed equalization (DDE), which is the more common equalization technique to refine an initial blind equalization. It can be seen that the performance is similar to DDE, obtaining a fine equalization. Moreover, the sampled-pdf approach does not require a decision and, therefore, it is not necessary to perform a previous phase identification.

6.3. Soft transition from blind to decision directed

In the following we will demonstrate the ability of this adaptation method to implement a soft transition between blind and decision directed-like equalization using (14) and (15). The method has been implemented for a 16 QAM modulation, and the compensation factor $F(\sigma)$ has been evaluated by means of a look-up table. The following channel [1, 2]

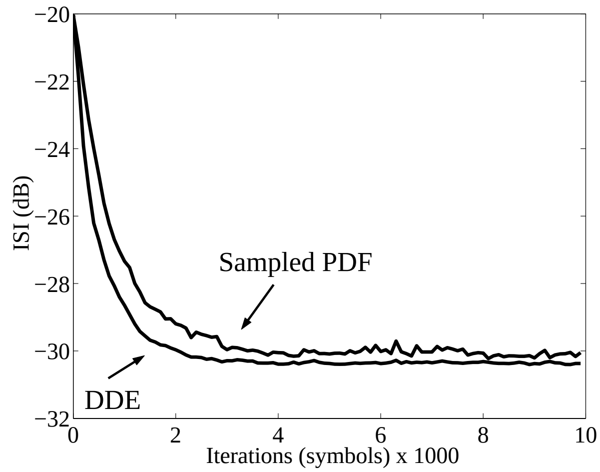

Fig. 4. Refinement with $\sigma = 1$ in $H_2(z)$ for a 16 QAM

has been employed to test this soft transition approach

$$H_3(z) = e^{j\theta} \frac{0.7 - z^{-1}}{1 - 0.7z^{-1}}. \tag{18}$$

In the first example, no phase rotation ($\theta = 0$) is considered. The following parameters have been used for the proposed method: $\mu = 0.02$, $\alpha = 0.99$, $a = 3.5$ and $b = -9.5$. E_1 has been initialized to start with a kernel size $\sigma_1 = 15$. Results are compared with those provided by the Benveniste-Goursat method [10]. A step size $\mu = 2e-4$ (the maximum for stable convergence) and the parameters recommended in [11] for digital TV channels are employed. Figure 5 compares the convergence of both methods.

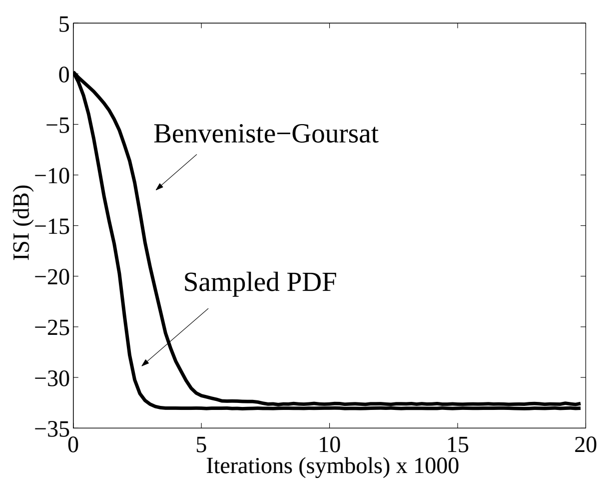

Fig. 5. Adaptive σ in $H_3(z)$ for 16 QAM with $\theta = 0$

The proposed approach is faster than the Benveniste-Goursat method and it provides an equivalent final accuracy. The advantage is more pronounced when the channel introduces some phase rotation. Figure 6 compares the convergence for $\theta = \pi/8$. Now the maximum step size providing stable convergence for the Benveniste-Goursat method is $\mu = 1e - 4$.

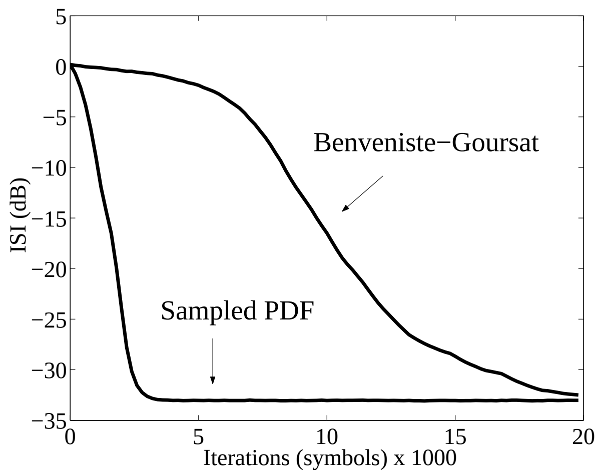

Fig. 6. Adaptive σ in $H_3(z)$ for 16 QAM with $\theta = \pi/8$

It can be seen that the performance of the Benveniste-Goursat method is clearly penalized by the phase rotation. However, the proposed approach is phase rotation invariant and, therefore, it provides the same results in both cases.

7. CONCLUSIONS AND FURTHER RESEARCH

A new method for blind deconvolution of multilevel modulations has been proposed. The proposed cost function is the squared error in the approximation of the pdf at certain sampling instants. The optimal sampling instants depend on the symbols of the constellation and on the kernel size of the Parzen window method. For a small kernel size the sampling points are specifically determined by the squared modulus of the symbols. However, for large kernel sizes a compensation factor must be introduced to ensure unity gain.

The proposed method has been tested for several channels in QAM modulations and it has demonstrated to be faster than CMA. Moreover, the algorithm is able to satisfactorily produce a soft transition between blind and decision directed operation. In this case, it outperforms the Benveniste-Goursat method. Therefore, this algorithm can be successfully employed in applications where both CMA and Benveniste-Goursat algorithms are already being used.

Further research is necessary to analyze the convergence behavior of the algorithm and to obtain analytical expressions for the compensation factor providing unity gain, but the experimental results have shown an excellent performance.

8. REFERENCES

[1] S. Haykin, *Blind Deconvolution*, Prentice Hall, Englewood, 1994.

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

[3] J. R. Treichler and B.G. Agee, “A new approach to multipath correction of constant modulus signals,” *IEEE Transactions on Acoustics, Speech, and Signal Proccesing*, vol. 31, pp. 349–372, Nov. 1983.

[4] D. N. Godard, “Self-recovering equalization and carrier tracking in two dimensional data communication systems,” *IEEE Transactions on Communications*, vol. 28, pp. 1867–1875, Nov. 1980.

[5] I. Santamaría, C. Pantaleón, L. Vielva, and J. C. Principe, “A fast algorithm for adaptive blind equalization using order-α Renyi’s entropy,” in *Proceedings of the International Conference on Acoustics, Speech, and Signal Proccesing*, Orlando, FL, USA, 2002, vol. III, pp. 2657–2660.

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

[7] J. Sala-Alvarez and G. Vazquez-Grau, “Stastistical reference criteria for adaptive signal processing in digital communications,” *IEEE Transactions on Signal Processing*, vol. 45, no. 1, pp. 14–31, Jan. 1997.

[8] T. Adali, X. Liao, and M. K. Sommez, “Conditional distribution learning with neural networks and its application to channel equalization,” *IEEE Transactions on Signal Processing*, vol. 45, no. 4, pp. 1051–1064, 1997.

[9] I. Santamaría, C. Pantaleón, L. Vielva, and J. C. Principe, “Adaptive blind equalization through quadratic pdf matching,” in *Proceedings of the European Signal Processing Conference*, Toulouse, France, Sept. 2002, vol. II, pp. 289–292.

[10] A. Benveniste and M. Goursat, “Blind equalizers,” *IEEE Transactions on Communications*, , no. 32, pp. 871–883, Aug. 1984.

[11] DIGISMATV, “SMATV reference channel model for digital TV,” Tech. Rep., DIGISMATV-NT-D1-031-HSA, 1995.