MATCHED PDF-BASED BLIND EQUALIZATION

*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, J. C. Principe

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

ABSTRACT

In this paper, a new blind equalization algorithm for multilevel modulations is proposed. It is based on maximizing the correlation between the probability density function (pdf) of the signal at the output of the equalizer and the desired pdf. The algorithm employs the Parzen window method to estimate the pdf of the squared modulus of the equalizer output. A stochastic gradient-based algorithm is used to maximize the correlation between this pdf and the pdf of the corresponding modulation. The proposed algorithm shows an excellent performance when compared with conventional adaptive blind algorithms, such as CMA, in quadrature amplitude modulation (QAM) schemes.

1. INTRODUCTION

Equalization is an important problem in digital high-speed communication systems and it has received a great amount of attention. Conventional equalization requires transmitting a training sequence that is known at the receiver. This sequence allows the adaptation of the equalizer parameters to minimize some error measurement (typically the mean square error) between the actual equalizer output and the desired response (the training sequence). When a linear filter is used to implement the equalizer, there are many adaptive algorithms that can be used to adapt the filter weights, for example the well known LMS [1].

When the transmission of a training sequence is not possible or practical, the problem at hand is named blind equalization. In this case, the only knowledge about the transmitted sequence is limited to its probabilistic or statistical properties. Blind equalization algorithms minimize a cost function that is able to indirectly extract the higher order statistics of the signal or the current level of ISI at the equalizer output [2]. Typically, the cost function is minimized by means of stochastic gradient algorithms. To this class of algorithms belong the Sato algorithm [3], which was the first blind technique for multilevel PAM signals, the Godard algorithms [4], or the Constant Modulus Algorithm (CMA) [5], a specific case of Godard algorithms, which is probably the most popular of blind equalization techniques. The main drawback of these algorithms when applied to communications is that they usually need a high number of data symbols to achieve convergence.

Several attempts have been made to improve the convergence speed of conventional blind techniques. Recently, Renyi's entropy has been introduced as a cost function for blind equalization of constant modulus signals [6]. This approach uses an efficient nonparametric estimator of this entropy measurement, based on the Parzen window method, to estimate the underlying pdf. Although this method provides excellent results for some channels, it fails to equalize other ones. The same authors have proposed, for constant modulus signals, a new method based on quadratic pdf matching [7]. In this case, the cost function is the integral of the squared difference between the estimated pdf and a target pdf. A simple Gaussian model is assumed for the target pdf.

In this paper, we propose a new cost function for multilevel modulations: the correlation between the equalizer output pdf and the desired one. The pdf is estimated using the well-known Parzen window technique, and a stochastic gradient algorithm has been developed to allow on-line implementation. This method has been applied to quadrature amplitude modulation (QAM) schemes, where it provides better results than the CMA algorithm.

2. THE NEW COST FUNCTION

For digital modulations, blind equalization can be stated 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 that usually is described by means of its discrete time complex coefficients h_k (assuming a FIR channel). In this case, the output channel can be obtained by

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

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

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

The blind equalizer will operate over the channel output in order to reduce, or ideally to completely remove, the intersymbol interference (ISI) introduced by the channel. In this approach we will use a linear equalizer implemented by means of a FIR filter; therefore, the equalizer output is given in this case by

$$y_k = \sum_{n=0}^{L_w-1} w_n x_{k-n} = \mathbf{w}^{\mathbf{T}}\mathbf{x}_k, \qquad (2)$$

where $\mathbf{w}$ is the vector of filter coefficients to be adapted by the blind equalization algorithm to minimize ISI. Usually, the blind algorithm makes use of *a priori* known statistics of the input signal. For instance, the Godard algorithms [4] propose to minimize the following cost function

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

where $R_p = E[|y_k|^{2p}]/E[|y_k|^p]$. CMA corresponds to the Godard algorithm for $p = 2$.

Information theory is an interesting alternative when the goal is to extract as much information as possible from the available data. The data distribution includes more information than the ratio R_p. In this case, it is possible to make use of the *a priori* knowledge of the probability density function (pdf) of $S^p = \{|s_k|^p\}$, and to try to minimize some distance between the actual pdf of the equalizer output and this desired pdf. There are several pdf divergence measures based on information theory: Kullback-Leibler's divergence, Bhattacharya distance, Chernoff distance or Renyi's divergence are some examples. In [8], the following cost function has been proposed in an application of information theory to clustering:

$$J(\mathbf{w}) = \int_{-\infty}^{+\infty} f_{Y^p}(z) f_{S^p}(z) dz, \qquad (4)$$

where $Y^p = \{|y_k|^p\}$, and $f_Z(z)$ denotes the pdf of Z at z. $J(\mathbf{w})$ is the correlation between both pdf's, and a maximum is obtained when the pdf of Y^p matches $f_{S^p}(z)$. Therefore, we have labeled this approach as *matched-pdf*.

It is necessary to note that, theoretically, the global maximum of (4) is obtained when $f_{Y^p}(z)$ is a delta located at the maximum of $f_{S^p}(z)$. In order to obtain the desired solution, some constraints have to be introduced. In this case, the underlying structure of the transmitted modulation and the pdf estimator (with a suitable σ) act as constraints, which prevent $f_{Y^p}(z)$ to converge to a delta function. This allows the algorithm to converge towards the desired solution, $f_{S^p}(z)$. Further work is necessary to formally state these constraints and their effects.

The Parzen window method is used to estimate the current pdf. Given a window of the L previous symbols, the estimate of the pdf $f_{Y^p}(z)$ at time k is

$$\hat{f}_{Y^p}(z) = \frac{1}{L}\sum_{i=0}^{L-1} K_{\sigma_o}(z - |y_{k-i}|^p). \qquad (5)$$

where $K_\sigma(x)$ is the Parzen window kernel of size σ. In this approach, Gaussian kernels with standard deviation σ are employed. For the sake of consistency, the target pdf must be the convolution of the original one with the kernel of the Parzen estimator we are using to estimate $f_{Y^p}(z)$, i.e.

$$\hat{f}_{S^p}(z) = \frac{1}{N_s}\sum_{i=0}^{N_s-1} K_{\sigma_o}(z - |s_i|^p), \qquad (6)$$

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

Finally, substituting (5) and (6) in (4), rearranging terms and taking into account that for Gaussian kernels

$$\int_{-\infty}^{+\infty} K_\sigma(y-C_1)K_\sigma(y-C_2)dy = K_{\sigma\sqrt{2}}(C_1-C_2), \qquad (7)$$

we obtain the following expression for the cost function

$$J(\mathbf{w}) = \frac{1}{LN_s}\sum_{i=1}^{N_s}\sum_{j=0}^{L-1} K_\sigma(|y_{k-j}|^p - |s_i|^p). \qquad (8)$$

For the sake of simplicity in the notation, we denote $\sigma_o\sqrt{2}$ as σ.

3. STOCHASTIC GRADIENT ALGORITHM

To minimize the cost function (8), a gradient descent technique will be used. In order to reduce the computational burden at each step, which is convenient to on-line implementations of the algorithm, we have considered a stochastic gradient approach using a window length $L = 1$. We will focus on $p = 2$, which is the more interesting case. Under these assumptions, the derivative of (8) with respect to the equalizer weights is given by

$$\frac{dJ(\mathbf{w}_k)}{d\mathbf{w}_k} = \frac{1}{N_s}\sum_{i=1}^{Ns} K'_\sigma(|y_k|^2 - |s_i|^2) y_k \mathbf{x}_k^*, \qquad (9)$$

where superscript * denotes complex conjugate and $K'_\sigma(x)$ is the derivative of the kernel function. Finally, the equalizer weights are adapted by

$$\mathbf{w}_{k+1} = \mathbf{w}_k + \mu_\sigma \frac{dJ(\mathbf{w}_k)}{d\mathbf{w}_k}. \qquad (10)$$

3.1. Implementation details

In (10), we have employed $\mu_\sigma = \mu\sigma^3$. The factor σ^3 has been introduced to compensate the term $1/\sigma^3$ that appears in $K'_\sigma(x)$ for Gaussian kernels. Moreover, taking into account that several symbols have the same module for the constellation of any modulation, the second term in the right side of (9) can be substituted by a weighted sum of all possible values of $|s_i|^2$. For instance, a 16QAM modulation has only 3 different values for $|s_i|^2$. This helps to reduce the computational burden of the method.

An important aspect of the algorithm is the selection of the kernel size σ. The kernel size determines the convergence speed and the accuracy of the final solution. For the sake of speed, a large kernel size is necessary, but we find the opposed requirement for the sake of accuracy. Convergence speed is the main requirement for equalizers working in communication systems. Typically, blind algorithms must operate until ISI is sufficiently reduced to open the eye of the constellation. At this moment, a switch to decision directed equalization is performed, providing an accurate final solution. Therefore, we will choice a large initial kernel size in order to reinforce convergence speed.

When using a large kernel size (a large amount of overlapping in $\hat{f}_{S^p}(y)$), the stochastic algorithm tends to slightly scale down the equalizer output. This behavior is due to the kernel overlapping, which affects the estimated pdf. In this case, the pdf at the output of the equalizer corresponding to the maximum of $J(\mathbf{w})$ is a scaled version of the desired pdf. This means that the matched pdf technique equalizes the channel up to a gain constant; the final gain can be easily compensated after convergence.

When a small kernel size is selected (a small amount of overlapping in $\hat{f}_{S^p}(y)$), the desired solution is a maximum of $J(\mathbf{w})$. Therefore, there is no scaling effect. Moreover, in this case a very accurate final solution can be obtained. Taking this into account, it is also possible to use the same algorithm switching only the kernel size as compared to switching to decision directed equalization as is done in CMA systems. We will show that this alternative provides similar results to decision directed equalization.

4. RESULTS

The proposed method has been tested in several different types of channels. In this section we will compare, for QAM signals, the performance of the proposed algorithm and the Constant Modulus Algorithm (CMA) [5] (Godard algorithm for $p = 2$), which is surely the most commonly employed blind algorithm in communication systems. The following figure of merit, which measures the intersymbol interference (ISI), will be used to compare the performance of the methods

$$ISI = 10\log_{10}\frac{\sum_n |\theta_n|^2 - \max_n |\theta_n|^2}{\max_n |\theta_n|^2}, \qquad (11)$$

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

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. For the equalizer, a filter with $L_w = 21$ taps and tap-centering initialization has been employed. A step size $\mu =$5e-5 and a kernel size $\sigma = 15$ have been selected for matched-pdf approach. This large kernel size has been selected because the main goal here is to maximize the convergence speed. Figure 1 compares the average results obtained in 100 Montecarlo trials with those obtained with the CMA algorithm using the step sizes $\mu =$1e-5 (the largest value for which all simulations converged).

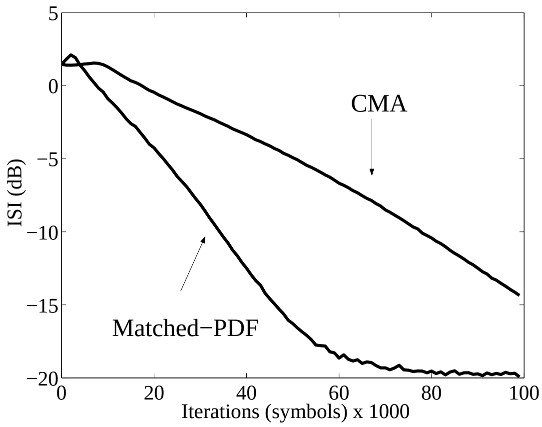

Fig. 1. Convergence for a 16 QAM system in channel $H_1(z)$

It can be seen how the matched-pdf approach converges faster than CMA. In this channel CMA shows a slow convergence, while the proposed approach is able to obtain a relevant improvement in convergence speed. The same behavior has been observed in several different channels. For instance, we have also tested the following nonminimum phase channel

$$H_2(z) = \frac{e^{j\frac{\pi}{4}}}{1.41}[0.4 - 0.6z^{-1} + 1.1z^{-2} - 0.5z^{-3} + 0.1z^{-4}].$$

In this case, $\mu = 3e-5$ and $\mu = 1e-4$ have been selected for CMA and matched-pdf respectively, which are the values that provide the fastest stable convergence in both cases. The convergence results are plotted in Figure 2.

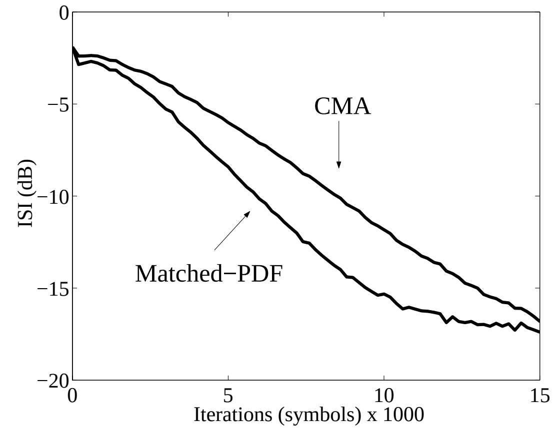

Fig. 2. Convergence for a 16 QAM system in channel $H_2(z)$

Again, the matched-pdf approach provides a better convergence than CMA. We want to remark that both methods have been tested over a large number of channels and we have not found a single case where CMA outperforms the proposed matched-pdf approach.

These results show that a large kernel size provides a fast convergence. However, to obtain a finer equalization, a small kernel size is more appropriate. Figure 3 shows the ISI evolution using $\sigma = 2$ for $H_2(z)$, after the initial convergence has been achieved using $\sigma = 15$. A scale correction has been performed before changing the kernel size. A simple measurement of the mean value of the current pdf has been employed to implement this correction. The matched-pdf performance is compared with decision directed equalization (DDE), which is the more common equalization technique to refine an initial blind equalization.

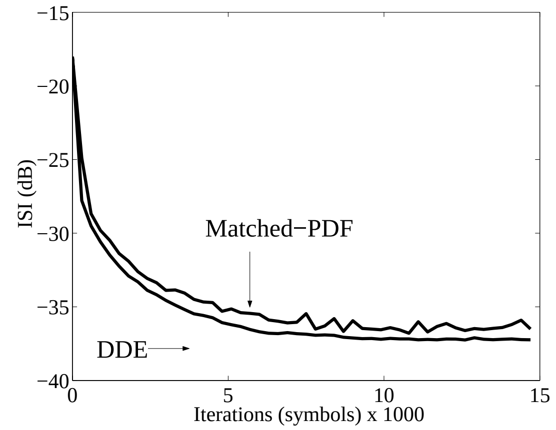

Fig. 3. Refinement with a small-kernel size in $H_2(z)$

We can see that the refinement obtained using matched-pdf is very similar to the DDE refinement. In this case, the small kernel size allows only the interaction of a constellation symbol $|s_i|^2$ with samples that are near enough in the squared modulus space. This limited interaction produces, in practice, a decision directed operation.

5. CONCLUSIONS

The correlation between the actual output pdf and the pdf of the underlying modulation has been introduced as a new cost function for blind equalization. In terms of convergence speed, the proposed method has shown better results than CMA in a large number of channels using 16QAM modulation.

On the other hand, the kernel size of the Parzen window method can be employed as a switching mechanism between blind and decision directed equalization. This strategy is able to provide the same refinement as conventional decision directed equalization.

6. REFERENCES

[1] B. Widrow and S. Stearms, *Adaptive signal processing*, Prentice Hall, Englewood Cliffs, NJ, 1985.

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

[3] Y. Sato, "A method of self-recovering equalization for multilevel amplitude modulation," *IEEE Transactions on Communications*, vol. COM-23, pp. 679–682, 1975.

[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] 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.

[6] 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.

[7] 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, 2002, vol. II, pp. 289–292.

[8] E. Gokcay and J. C. Principe, "Information theoretic clustering," *IEEE Trans. on Pattern Analysis and Machine Inteligence*, vol. 24, no. 2, pp. 158–171, 2002.