A Bilinear Approach to Constant Modulus Blind Equalization

Constantinos B. Papadias and Dirk T. M. Slock

Institut EURECOM, 2229 route des Crêtes,
B.P. 193, 06904, Sophia Antipolis Cedex, FRANCE

e-mail: papadias@eurecom.fr, slock@eurecom.fr

Abstract

We consider the problem of blind equalization of a constant modulus signal that is received in the presence of Inter-Symbol-Interference (ISI) and additive noise. A well-known class of adaptive algorithms for this problem is the so-called Godard family of blind equalizers[1], including among others the CMA[2] and SATO[3] algorithms. These algorithms are known for their ability in general to open the eye of a communications channel and for their low computational complexity. However, a common disadvantage of all algorithms of this class is that they might exhibit ill-convergence if not properly initialized, due to the non-convex form of their cost function. In this paper we present a different approach to the problem, namely, a *bilinear* approach in order to construct a convex cost function with a unique minimum point. After presenting the formulation for this approach, we show that in the case of an exactly invertible noiseless channel the optimal solution that completely opens the communication channel's eye may always be attained, regardless of the initial equalizer's setting. This implies that equalization may be also achieved for a noisy FIR channel, provided that the equalizer is long enough to approximate adequately the channel's inverse impulse response. Computer simulations are provided to show the validity of our theoretical arguments.

1 Introduction

Consider the problem of blind equalization of a constant modulus signal as depicted in figure 1, where a baseband representation of a communications system is shown. The samples of the emitted signal

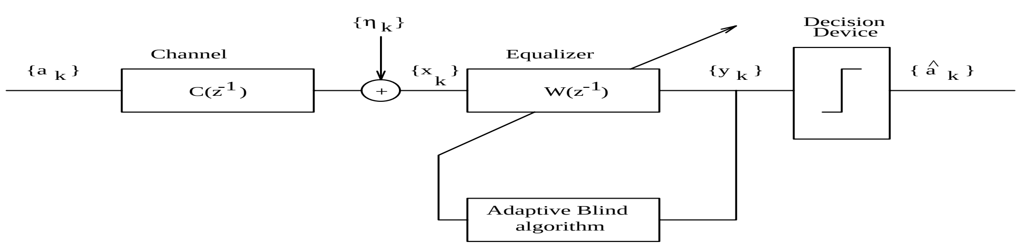

Figure 1: A typical blind equalization scheme

$\{a_k\}$ (with constant modulus) are transmited through a linear noisy channel resulting in the received sequence of samples $\{x_k\}$ that are both corrupted by Inter-Symbol-Interference (ISI) and additive noise. The task of the blind equalizer is to match the inverse of the linear channel's impulse response in order to cancel the ISI and result in a correct retrieval of the emitted symbols $\{a_k\}$ based only on statistical information about the emitted symbols. A very popular class of algorithms that try to cancel ISI by exploiting the constant-modulus property of the constellation is the so-called Godard family of blind equalizers, which includes among others the well known Constant Modulus Algorithm (CMA). This algorithm minimizes the following cost function:

$$J_p(W) = \frac{1}{2p} E(|y|^p - R_p)^2 \qquad , \; p \in 1, 2, \dots \; , \tag{1}$$

where E denotes statistical expectation and R_p is a constant scalar called *dispersion constant* and defined as $R_p = \frac{E|a_k|^{2p}}{E|a_k|^p}$. The corresponding stochastic gradient algorithm is given by:

$$W_{k+1} = W_k - \mu X_k y_k |y_k|^{p-2}(|y_k|^p - R_p) \ . \tag{2}$$

In the above notation, W_k is a column vector containing the equalizer's setting at time instant k, $X_k = [x_k \ x_{k-1} \cdots x_{k-N+1}]^H$ and y_k denotes the equalizer's output at time instant k and may be written as $y_k = X_k^H W_k$ (^{H} denotes complex conjugate transposition). The well known CMA 2-2 algorithm is a special case of (2) for $p = 2$ and has been shown to be able in general to open the communication system's eye. A main drawback however of all Godard equalizers is that they might converge to undesired solutions if not properly initialized, due to the non-convex form of their cost function which has a number of local minima apart from its global minimum point. For example, the equilibria points for the algorithm (2) (p=2) are given by the following equation:

$$E((|y_k|^2 - R_2) y_k X_k) = 0 \ . \tag{3}$$

Equation (3) is a system of N nonlinear equations for the coefficients of the equalizer filter W. The highly non-linear character of this system of equations results in a plenitude of solutions, some of which are minima of the cost function $J_2(W)$. This is why false minima possibilities exist for the CMA algorithm. We will now propose a different approach to the problem of figure 1 in order to avoid the problems arising from the non-convexity of the cost function in (1).

2 A bilinear approach

2.1 Formulation

We now suppose that the channel is a possibly non-minimum phase linear filter with an inverse impulse response of order N. We also assume that the emitted symbols $\{a_k\}$ may equally likely take on the binary values 1 and -1 (2-PAM modulation) (the complex(QAM) case may be handled equally well). Consider now the following regression vector:

$$\mathcal{X}_k = [x_k^2 \ \ 2x_k x_{k-1} \ \cdots \ x_{k-1}^2 \ \ 2x_{k-1}x_{k-2} \cdots \ x_{k-N+1}^2]^T \ , \tag{4}$$

that contains all possible product combinations (square terms and cross-terms) of the received samples $x_k, \cdots, x_{k-N+1}$ with the convention that a multiplicative factor of 2 is used for products of different samples (cross-terms). This vector has $\sum_{i=1}^{N} i = \frac{N\times(N+1)}{2}$ entries (which are actually the terms in the expansion of the quantity $(\sum_{i=0}^{N-1} x_{k-i})^2$). Consider also the following vector θ with $\frac{N\times(N+1)}{2}$ entries:

$$\theta = [\theta^{(0)} \ \ \theta^{(1)} \ \cdots \ \theta^{(\frac{N\times(N+1)}{2}-1)}]^T \ . \tag{5}$$

This vector will denote the impulse response of a linear filter through which passes the regression vector $\mathcal{X}_k$ and its value at time instant k will be denoted by θ_k. The output of this filter at time instant k will be denoted by z_k and may be expressed as $z_k = \mathcal{X}_k^T \theta_k$. Our aim will be to force this filter's output to the positive constant 1 by penalizing its deviations from this constant in a square-sense:

$$\min_{\theta} E(z-1)^2 \ . \tag{6}$$

The philosophy behind this criterion is that, in a way, we take the "square" of the received signal and then convolve it with a filter in order to produce an output equal to the square of the emitted signal's modulus. This must be seen in contrast to traditional equalizers, where the received signal is first passed through an equalizer whose *output's* square is forced to match the emitted signal's squared modulus. As will be shown, this "interchange" of non-linearity and equalizer still allows for an identification of the channel's inverse impulse response. Moreover, it will provide a unique solution thus avoiding the problem of false-minima.

The criterion in (6) corresponds to a quadratic cost function and has the following unique solution:

$$\tilde{\theta} = \{E(\mathcal{X}\mathcal{X}^T)\}^{-1}E(\mathcal{X}) \ , \tag{7}$$

provided that the inverse of the "covariance" matrix $E(\mathcal{X}\mathcal{X}^T)$ exists. On the other hand, one solution to the problem (5) is the following:

$$\bar{\theta} = [w_0^2 \ \ w_0 w_1 \ \ \cdots w_1^2 \ \ w_1 w_2 \ \cdots \ \ \cdots \ w_{N-1}^2]^T \ , \tag{8}$$

where $\tilde{W} = [w_0 \ w_1 \cdots w_{N-1}]^T$ represents the impulse response of the true inverse of the transmission channel. This happens because at each time instant, $\mathcal{X}_k^T \bar{\theta} = (X_k^T W)^2 = a_k^2 = 1$ (in the absence of noise or approximation error) and thus with $\theta = \bar{\theta}$ the cost function $E(z-1)^2$ achieves its minimum value 0:

$$E(\mathcal{X}^T \bar{\theta} - 1)^2 = 0 \ . \tag{9}$$

The above considerations lead to the following lemma:

Lemma: When the impulse response of the inverse of the transmission channel is of length N, no additive noise is present and the matrix $E(\mathcal{X}\mathcal{X}^T)$ is invertible, then the problem (6) has the unique solution (8).

The calculation of this solution automatically implies the identification of the channel's inverse impulse response, since information about both the magnitude and the sign of the elements of $\tilde{W}$ is contained in $\bar{\theta}$. This means that, from an identification point of view, the non-minimum phase transmission channel may be determined (under some idealized conditions with perfect accuracy) by (7). Intuitively, this is not an astonishing result, since the quantity $E(\mathcal{X}\mathcal{X}^T)^{-1}E(\mathcal{X})$ contains statistics of orders higher than 2 (HOS) that are necessary, as is known, for the identification of a non-minimum phase system.

The above lemma implies also that, in the case of a FIR linear channel, an equalizer long enough to adequately approximate the channel's inverse will still converge to an acceptable setting that opens the eye of the system.

We will in the sequel examine how the solution in (7) may be calculated and then how one determines from this solution the impulse response $\tilde{W}$.

2.2 Determination of the channel's inverse impulse response

One may now distinguish between two different general methodologies in order to estimate the channel's inverse: either use an adaptive algorithm to recursively compute at each step an update for θ_k and then determine the corresponding channel setting for this step, or use some kind of estimator to calculate the quantity $E(\mathcal{X}\mathcal{X}^T)^{-1}E(\mathcal{X})$ and then determine the corresponding channel based on this estimation only once (batch processing).
An adaptive equalization setup based on the above principle is depicted in figure 2. In this figure, the $\frac{N\times(N-1)}{2}$ entries of the equalizer θ are "distributed" in N linear equalizers of respective lengths $N, \ldots, 1$, i.e. a multichannel structure is used (however this structure is equivalent to using only one equalizer θ with $\frac{N\times(N-1)}{2}$ entries). The i^{th} equalizer has as input the stream $x_k x_{k-i+1}$. The entries of the regression vector in (4) are calculated by using a tapped-delay line and some multiplication operators. The output of the bank of N equalizers at time instant k, z_k, is then substracted from the constant 1 thus creating the error ϵ_k which feeds the adaptive algorithm that tries to minimize the squared error by adapting the equalizers' setting. This can be done by any classical adaptive filtering algorithm, like e.g. an LMS or an RLS algorithm or any other more sophisticated adaptive filtering algorithm. For example, an RLS

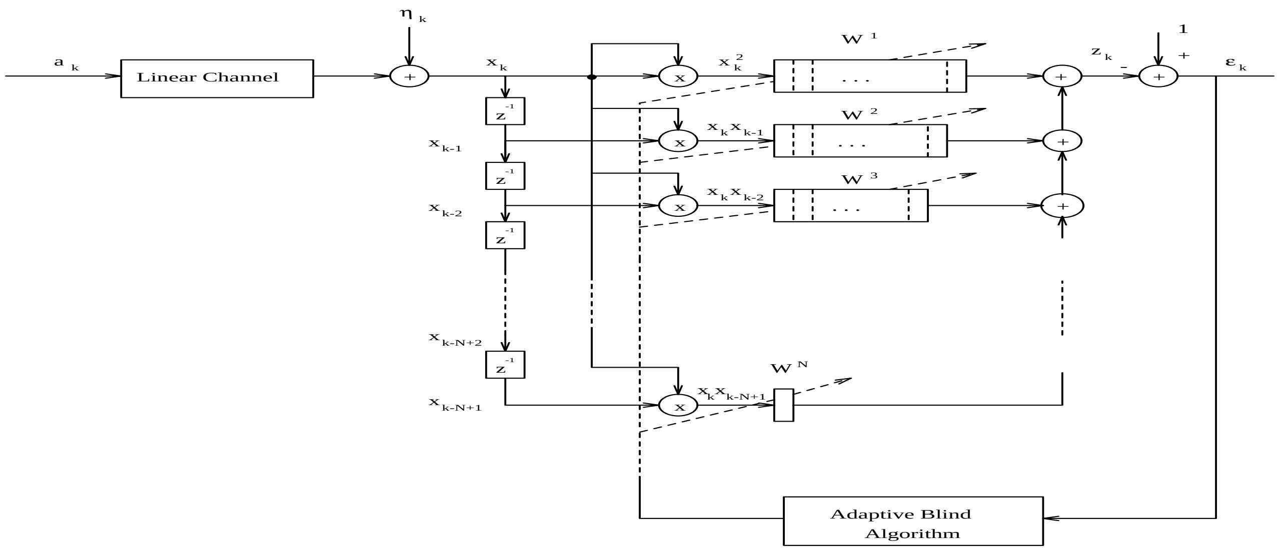

Figure 2: A bilinear blind equalization setup

algorithm to recursively compute the solution in (7) would essentially perform the following operations at each time instant:

$$\begin{aligned} \epsilon_k &= 1 - \mathcal{X}_k^T \theta_k \\ R_k^{-1} &= \lambda^{-1} R_{k-1}^{-1} - \lambda^{-1} R_{k-1}^{-1} \mathcal{X}_k (1 + \mathcal{X}_k^T \lambda^{-1} R_{k-1}^{-1} \mathcal{X}_k)^{-1} \mathcal{X}_k^T \lambda^{-1} R_{k-1}^{-1} \\ \theta_{k+1} &= \theta_k + R_k^{-1} \mathcal{X}_k \epsilon_k \ . \end{aligned} \tag{10}$$

In the same way, an LMS algorithm would be as follows:

$$\begin{aligned} \epsilon_k &= 1 - \mathcal{X}_k^T \theta_k \\ \theta_{k+1} &= \theta_k + \mu \mathcal{X}_k \epsilon_k \ . \end{aligned} \tag{11}$$

Independently of our choice of algorithm, it will converge after a number of operations to its unique minimum point given in (7). It is then our task to determine $\widetilde{W}$ from $\widetilde{\theta}$ (of course this can also be done at each iteration in order to gradually open the system's eye). A way to do this is the following: we collect all the $\frac{N \times (N+1)}{2}$ elements of vector θ into a symmetric $N \times N$ matrix Θ as following:

$$\Theta = \begin{bmatrix} \theta^{(0)} & \theta^{(1)} & \dots & \theta^{(N)} \\ \theta^{(1)} & \theta^{(N+1)} & \dots & \theta^{(2N-1)} \\ . & . & \dots & . \\ . & . & \dots & . \\ \theta^{(N)} & \theta^{(2N-1)} & \dots & \theta^{(\frac{N \times (N+1)}{2} - 1)} \end{bmatrix} \ . \tag{12}$$

If θ has attained exactly its optimal value predicted in (8), the above matrix Θ will be as follows:

$$\bar{\Theta} = \begin{bmatrix} w_0^2 & w_0 w_1 & \cdots & w_0 w_{N-1} \\ w_0 w_1 & w_1^2 & \cdots & w_1 w_{N-1} \\ . & . & \cdots & . \\ . & . & \cdots & . \\ w_0 w_{N-1} & w_1 w_{N-1} & \cdots & w_{N-1}^2 \end{bmatrix} , \tag{13}$$

which is a rank-1 matrix ($\bar{\Theta} = \widetilde{W} \widetilde{W}^T$). Then $\widetilde{W}$, the optimal equalizer setting, may be directly computed as:

$$\widetilde{W} = \sqrt{\lambda_{max}} \, V_{max} \ , \tag{14}$$

where λ_{max} and V_{max} denote the largest eigenvalue and the corresponding eigenvector of matrix $\bar{\Theta}$, respectively (all other eigenvalues should ideally equal zero). Therefore, we may directly determine from θ the impulse response of the channel's inverse by first forming the matrix Θ and then computing its largest eigenvalue and the corresponding eigenvector.

3 Over-parameterized equalizer: a special case

A case that merits a special consideration is the case of an over-parameterized equalizer w.r.t. the channel's impulse response (even though this case is not a very realistic one). Such a situation may arise, e.g. when the channel is a AR($N-1$) filter and the equalizer θ has $\frac{M\times(M+1)}{2}$ elements ($M > N$), i.e. more than the strictly needed ($\frac{N\times(N+1)}{2}$) in order to perfectly match the channel's inverse impulse response. In this case, there exist more than one solution that perfectly identify the channel's inverse, corresponding to shifted versions of the ideal setting. For example, suppose an AR(1) channel $C = [c_0\ c_1]$ and a MA equalizer of 6 entries (instead of 3): $\theta = [\theta^{(0)}\ \theta^{(1)}\ \theta^{(2)}\ \theta^{(3)}\ \theta^{(4)}\ \theta^{(5)}]^T$. It is obvious that both settings $\theta^1 = [0\ 0\ 0\ c_0^2\ c_0c_1\ c_1^2]^T$ and $\theta^2 = [c_0^2\ c_0c_1\ 0\ c_1^2\ 0\ 0]^T$ perfectly open the system's eye, since they correspond to the respective settings $W^a = [0\ c_0\ c_1]^T$ and $W^b = [c_0\ c_1\ 0]^T$. In fact, any vector θ of the form $\theta = \alpha\theta^1 + \beta\theta^2$ with $\alpha + \beta = 1$ satisfies the following equation:

$$E(\mathcal{X}\mathcal{X}^T)\theta = E(\mathcal{X}) \ , \tag{15}$$

since $E(\mathcal{X}^T\theta^i) \equiv 1$, $i = 1, 2$. This means that, except for the cases $[\alpha\ \beta] = [1\ 0]$ and $[\alpha\ \beta] = [0\ 1]$, the solution for θ found by the algorithm will not directly correspond to the true channel setting but to a weighted sum of two shifted versions of it. This generalizes for any value of $M > N$ to a weighted sum of $M - N$ shifted versions of the true channel setting. We now present a method to face this problem. Consider the eigenvalue decomposition of matrix Θ as follows:

$$\Theta = \sum_{i=1}^{N} \lambda_i V_i V_i^T \ , \tag{16}$$

where the eigenvectors V_i are considered to be in a descending order according to the absolute values of their eigenvalues. Consider also the following Toeplitz matrix :

$$\mathcal{W} = \begin{bmatrix} w_0 & w_{-1} & \cdots & w_{-n} \\ w_1 & w_0 & \ddots & \vdots \\ \vdots & \vdots & \ddots & w_0 \\ \vdots & \vdots & \ddots & \vdots \\ w_{N-1} & w_{N-2} & \cdots & w_{N-1-n} \end{bmatrix} \ , \tag{17}$$

where the elements $w_{-1}, \cdots, w_{-n}$, play the role of some additional coefficients in the impulse response of the channel's inverse and n is a small integer ($n < N$ and ideally $n = M - N$). We now consider the following problem:

$$\min_{Q, \{w_{-n}, \cdots, w_{N-1}\}} \|\mathcal{W} - \mathcal{V}Q\|_F^2 \ , \tag{18}$$

where $\|.\|_F$ denotes the Frobenius norm ($\|A\|_F^2 = tr(A^T A)$), $Q \in \mathcal{R}^{(n+1)\times(n+1)}$ and $\mathcal{V}$ is a matrix containing the $n+1$ first eigenvectors of Θ: $\mathcal{V} = [V_1 \cdots V_{n+1}]$. It can be shown that the solution to the above problem (18) (a vector of $n+N$ entries) is the eigenvector corresponding to the largest eigenvalue of the following matrix:

$$\Lambda = \sum_{i=1}^{n+1} \Omega_i \Omega_i^T \ , \qquad \Omega_i = [0^T_{(n-i+1)\times(n+1)}\ \ \mathcal{V}^T\ \ 0^T_{(i-1)\times(n+1)}]^T \ . \tag{19}$$

This method has been proven to work perfectly in over-determined cases that we examined via computer simulations.

4 Simulations

The methods we proposed above have been tested via computer simulations that have verified the validity of our arguments. In all noiseless cases of AR channels and MA equalizers exactly parameterized, perfect identification has been observed, as expected, independently of the algorithms' initializations. Such an example may be seen in figure (3a), which refers to an AR(1) noiseless channel and an equalizer $W = [w_0\ w_1]^T$ corresponding to $\theta = [\theta^{(0)}\ \theta^{(1)}\ \theta^{(2)}]^T$. The employed algorithm is (11) and the starting and ending points of 40 different initializations on a circle of radius 2 are shown. One may see the global convergence property of the algorithm. The same experiment has been carried out with CMA and the result is shown in figure (3b) which shows the existence of local minima. A more realistic simulation is shown in figures (3c) and (3d) where one may see the evolution of the closed-eye measure of a linear noisy (SNR=30 dB) communications system using the algorithm in (10) ($\lambda = 1$) and CMA, respectively. The channel's impulse response is [1 0.6 0.36] and an equalizer W of 8 taps is used (28 taps for θ). One may see how the opening of the system's eye may be achieved for 2 different initializations by using a bilinear algorithm, while CMA gets trapped by a local minimum for one of these initializations.

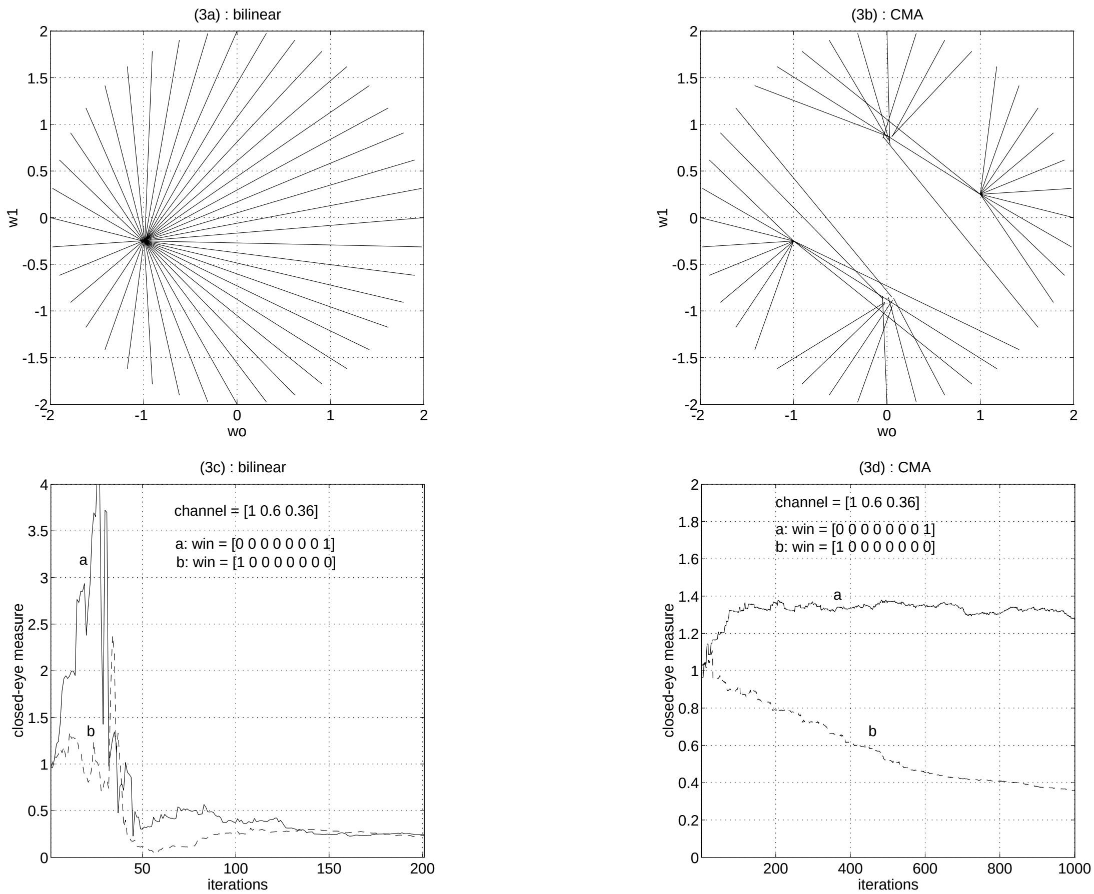

Figure 3: Computer simulations results

5 Conclusions

We have proposed a new approach to the problem of blind equalization of a constant modulus signal that is corrupted by Inter-Symbol-Interference and additive noise. This approach introduces a regression vector that contains bilinear terms of the received distorted sequence and uses this vector in order to identify the channel's inverse impulse response, hence the term *bilinear*. The channel's inverse impulse response is found in two steps: first a linear filter through which the bilinear regressor passes is computed so as to produce a constant output and then this setting is used in order to find the desired impulse response. This approach leads to globally convergent algorithms, since it minimizes a convex (quadratic) cost function with a unique minimum point. Algorithmic aspects such as complexity, speed and robustness of different schemes are the object of current work.

REFERENCES

[1] D. N. Godard, "Self-Recovering Equalization and Carrier Tracking in Two-Dimensional Data Communications Systems," *IEEE Trans. Commun.*, vol. COM-28, pp. 1867-1875, Nov. 1980.

[2] J. R. Treichler and B. G. Agee, "A new Approach to Multipath Correction of Constant Modulus Signals," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-31, pp. 459-472, Apr. 1983.

[3] Y. Sato, "A Method of Self-Recovering Equalization for Multilevel Amplitude Modulation Systems," *IEEE Trans. on Commun.*, vol. COM-23, pp. 679-682, June 1975.