A Class of Channels Resulting in Ill-Convergence for CMA in Decision Feedback Equalizers

Aline Neves and Cristiano Panazio, *Member, IEEE*

***Abstract*—This paper analyzes the convergence of the constant modulus algorithm (CMA) in a decision feedback equalizer using only a feedback filter. Several works had already observed that the CMA presented a better performance than decision directed algorithm in the adaptation of the decision feedback equalizer, but theoretical analysis always showed to be difficult specially due to the analytical difficulties presented by the constant modulus criterion. In this paper, we surmount such obstacle by using a recent result concerning the CM analysis, first obtained in a linear finite impulse response context with the objective of comparing its solutions to the ones obtained through the Wiener criterion. The theoretical analysis presented here confirms the robustness of the CMA when applied to the adaptation of the decision feedback equalizer and also defines a class of channels for which the algorithm will suffer from ill-convergence when initialized at the origin.**

***Index Terms*—Blind equalization, constant modulus algorithm, decision directed algorithm, decision feedback equalization, ill-convergence.**

I. Introduction

TIME dispersive channels generate intersymbol interference that hinders the performance of digital communication systems. To overcome such issue, we may use an equalizer. Linear equalization provides a low computational complexity solution, but suffers from noise-enhancement, presenting a performance that is not satisfactory for channels with spectral nulls [1]. On the other hand, maximum-likelihood sequence estimation [2]–[4] gives the optimal bit-error rate at the cost of a huge complexity, specially for long channels and higher order modulations. A good compromise between both solutions is provided by a nonlinear approach namely decision feedback equalization [5], [6]. Such solution is basically an infinite impulse response filter with a nonlinear decision device in the feedback loop. The addition of this device practically eliminates the noise-enhancement, even in channels with spectral nulls.

The most used and studied blind algorithm in the decision feedback equalizer (DFE) context is the decision directed algorithm (DD-DFE). This algorithm is known to converge well, attaining the optimum solution, when the feedback filter is initialized at the origin and the channel is not severe, i.e., in an opened-eye situation. In severe channels, however, the algorithm may converge to local minima that do not reduce intersymbol interference. In this case, a good initialization is crucial to attain a good solution. Usually, decision directed algorithms are used after a first period of trained adaptation, which sufficiently reduces intersymbol interference leading to an opened-eye situation.

In a linear equalization structure, where the equalizer is given by a finite impulse response (FIR) filter, another blind criterion is known to perform better than decision directed: the constant modulus (CM) criterion [7], [8]. This criterion presents fewer local minima and is able to considerably reduce intersymbol interference even in a closed-eye situation. Since it presents a better performance when used in linear structures, the extension of its use to nonlinear structures such as the decision feedback equalizer was already expected. Several works [9]–[11] tested the CM algorithm (CMA-DFE) in this context and its better performance was observed through simulations. Strategies combining both algorithms, starting the adaptation with an infinite impulse response filter adapted by CMA and after switching to a DD-DFE were also tested [10], [12]. Due to its analytical difficulties, however, no theoretical convergence analysis on the CMA-DFE was done hitherto.

Unlike CMA-DFE, DD-DFE has been extensively studied in the literature [13], [14]. In [14], the authors studied the DD-DFE convergence properties classifying the local minima into two types: the ones that resulted from the decision feedback equalizer structure, and the ones resulting from the adaptive algorithm itself. The first type of minima does not depend on the algorithm being used and was carefully studied by the authors in [14]. The second type of minima was carefully analyzed in [13]. In this work, the authors were able to define a class of channels for which DD-DFE converges to a bad local minimum when the feedback filter is initialized at the origin. Such initialization is interesting when no *a priori* information on the channel is known.

In this paper, our objective is to extend the results obtained in [13] to the CMA-DFE. The first step in this direction is to consider the same framework as [13], where the decision feedback equalizer is given only by the feedback filter. This condition also restrain the channels being studied. We will consider that they do not present precursor coefficients and that no noise is added. Even though such assumptions are very limiting for practical purposes, they have been considered by several authors in order to render the decision feedback equalizer analysis, that is not trivial, more feasible [11], [14]. It should be noted that the performance of a decision feedback equalizer is largely altered when considering both filters, feedforward and feedback. Since, in this case, a theoretical analysis would be impractical,

Manuscript received January 04, 2010; accepted July 27, 2010. Date of publication August 16, 2010; date of current version October 13, 2010. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Alper Tunga Erdogan.

A. Neves is with the Centro de Engenharia, Modelagem e Ciências Sociais Aplicadas—Universidade Federal do ABC 09210-170 Santo André, São Paulo, Brazil (e-mail: aline.neves@ufabc.edu.br).

C. Panazio is with the Escola Politécnica of the University of São Paulo, 05508-900 São Paulo, São Paulo (e-mail: cpanazio@lcs.poli.usp.br).

Digital Object Identifier 10.1109/TSP.2010.2065225

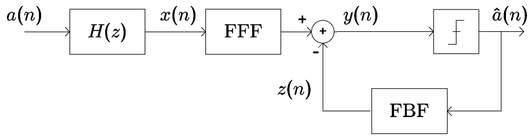

Fig. 1. Decision feedback equalizer.

comments on the behavior of a decision feedback equalizer with both filters, adapted by the CMA, may be found at the end of this paper.

The analytical difficulties that had hindered the theoretical development towards the understanding of the CMA-DFE convergence performance may be bypassed by a recent result obtained in [15] and [16]. Several other works in the literature also compare Wiener filters with constant modulus solutions [17]–[20], arriving at mean square error performance bounds for the constant modulus solutions and showing that such solutions are collinear with Wiener receivers. However, in [15] and [16], the authors compare the CM minima with the ones obtained using the classical supervised Wiener criterion using a very clever approximation that enables the analytical obtention of the CM minima. Such analysis consider a linear FIR equalizer, but it can be directly extended to the decision feedback equalizer scenario.

This paper is organized as follows. Section II presents the system model to be considered throughout the work. Section III defines the DD-DFE, while Section IV presents the CMA-DFE, analyzing the constant modulus criterion in order to determine the minima locations in the decision feedback equalizer scenario. Section V reviews the results presented in [13] for three-tap channels, analyzing the DD-DFE error surface and extending the results to CMA-DFE. Section VI generalizes the obtained result for a $N+1$-order channel. Finally, Section VII concludes this paper.

II. System Model

The most general form of a DFE may be represented as shown in Fig. 1. Symbols $a(n)$ are transmitted through a channel $H(z)$. The received data with noise $x(n)$ is first filtered by a FIR feedforward filter. The output of such a filter is subtracted from the output of the feedback filter $z(n)$ which is a linear combination of previously taken decisions. Such operation leads to $y(n)$. Assuming correct decisions, the feedback filter output may be viewed as the amount of postcursor intersymbol interference present in the output of the feedforward filter.

Since our objective is to extend the results obtained by [13], we will first place ourselves in the same problem context. In this sense, the first important assumption to be made can be stated as follows.

Assumption 1: The channel has a finite impulse response and there is no precursor, i.e., the leading tap dominates. Therefore, we can define the vector $\mathbf{h} = [h_0\, h_1\, \ldots h_N]^T$, composed by the $N+1$ elements of the channel impulse response, where $h_0 = 1$ and $|h_i| \leq 1$ for $i = 1, 2, \ldots, N$. We also assume that no noise is added.

It is clear that Assumption 1 is very limiting for practical channels, where usually precursors and noise are present. As a consequence of this assumption, the feedforward filter is not useful and can be discarded, what enables us to restrict our analysis to the feedback filter and to the local minima associated with error propagation. Several works [13], [14] consider this same restriction, allowing the analytical tractability of the decision feedback equalizer.

The restriction $|h_i| \leq 1$ for $i = 1, 2, \ldots, N$ derives from the fact that we are concerned with minima resulting from the adaptive algorithm itself. When the channel presents a precursor, decision feedback equalizer may lead to what can be called as delay-type minima, which were studied in detail by [14]. Thus, such a restriction avoids delay-type minima. In addition, when using a feedforward filter, which mitigates the effect of the precursor, the channel-feedforward filter response seen by the feedback filter will often fall in the class of channels defined by Assumption 1 [9].

Furthermore, we will consider channels with real valued coefficients, even though the extension to the complex case is direct.

Under Assumption 1, the decision feedback equalizer structure may be described mathematically as follows:

$$y(n) = (x(n) - z(n)) = \left(\boldsymbol{a}_n^T \mathbf{h} - \mathbf{w}_n^T \hat{\boldsymbol{a}}_n\right) \tag{1}$$

where $x(n)$ is the received signal, $z(n)$ is the feedback filter output signal, $\boldsymbol{a}_n = [a(n)\, a(n-1) \ldots a(n-N)]^T$ is the vector of transmitted symbols, $\mathbf{w}_n = [w_1(n) \ldots w_N(n)]^T$ is the feedback filter tap vector and $\hat{\boldsymbol{a}}_n = [\hat{a}(n-1)\, \hat{a}(n-2) \ldots \hat{a}(n-N)]^T$ is the feedback filter input, given by the past decided symbols. Before defining the decision device function, we need to state our second assumption, also with the objective of rendering the convergence analysis more feasible.

Assumption 2: The transmitted symbols are binary phase shift keying (BPSK), belonging to the alphabet $\{-1, +1\}$, and are independent and identically distributed.

Therefore, the decision device nonlinear function is given by

$$\hat{a}(n) = \text{sign}(y(n)) = \text{sign}(x(n) - z(n)). \tag{2}$$

In addition, in order to guarantee that perfect equalization is possible, see Assumption 3 below.

Assumption 3: The feedback filter length matches the channel postcursor, that is, the feedback filter length is equal to N.

Having defined the system model in which this work is inserted, let us discuss the two most used blind algorithms for the feedback filter adaptation.

III. Decision-Directed Decision Feedback Equalizer

The decision-directed criterion is based on the minimization of the mean square error between the decided symbols and the equalizer output. In the decision feedback equalizer context, its cost function may be written as

$$J_{\text{DD}} = E\{(y(n) - \hat{a}(n))^2\}. \tag{3}$$

The minimization of (3) leads to the well known least mean squares-type decision feedback equalizer algorithm named decision directed algorithm (DD-DFE):

$$\begin{aligned} e(n) &= (y(n) - \hat{a}(n)) \\ \mathbf{w}_{n+1} &= \mathbf{w}_n + \mu e(n) \hat{\boldsymbol{a}}_n \end{aligned} \tag{4}$$

where μ is the step size.

Analyzing (3) more carefully, and using (1), we may rewrite $J_{\rm DD}$ as

$$J_{\rm DD} = E\left\{\left(\boldsymbol{a}_n^T\mathbf{h} - \mathbf{w}_n^T\hat{\boldsymbol{a}}_n - \hat{a}(n)\right)^2\right\}. \quad (5)$$

Differentiating (5) with respect to $\mathbf{w}_n$ and setting it equal to zero, we arrive at [13]

$$\mathbf{w}_{\rm DD}^* = E\left\{\hat{\boldsymbol{a}}_n\hat{\boldsymbol{a}}_n^T\right\}^{-1}\left[E\left\{\hat{\boldsymbol{a}}_n\boldsymbol{a}_n^T\right\}\mathbf{h} - E\{\hat{\boldsymbol{a}}_n\hat{a}(n)\}\right] \quad (6)$$

which defines the optimum value, $\mathbf{w}_{\rm DD}^*$, of the feedback filter coefficients. The statistics present in (6) depend on the region of the error surface, denoted by $\mathcal{P}$, in which the filter taps are placed. This topic will be discussed in detail in Section V. Defining $A(\mathcal{P}) = E\{\hat{\boldsymbol{a}}_n\hat{\boldsymbol{a}}_n^T\}$, $C(\mathcal{P}) = E\{\hat{\boldsymbol{a}}_n\boldsymbol{a}_n^T\}$ and $T(\mathcal{P}) = E\{\hat{\boldsymbol{a}}_n\hat{a}(n)\}$, (6) may be rewritten as

$$\mathbf{w}_{\rm DD}^* = A(\mathcal{P})^{-1}[C(\mathcal{P})\mathbf{h} - T(\mathcal{P})]. \quad (7)$$

IV. Constant Modulus Algorithm Based Decision Feedback Equalizer

The constant modulus criterion penalizes deviations of the equalizer output from a constant modulus:

$$J_{\rm CM} = E\{((y(n))^2 - R)^2\} \quad (8)$$

where $R = (E\{(a(n))^4\})/(E\{(a(n))^2\})$. Following Assumption 2, R will be equal to one. The resulting gradient descent algorithm, known as CMA, is given by

$$\mathbf{w}_{n+1} = \mathbf{w}_n + \mu y(n)((y(n))^2 - R)\hat{\boldsymbol{a}}_n. \quad (9)$$

It is important, for the rest of our analysis, to be able to analytically obtain the position of the optimum solution $\mathbf{w}_{\rm CM}^*$ resulting from the minimization of (8). In fact, works like [13] and [14] did not include CMA in their analysis because of the difficulty in achieving such result.

Differentiating (8) with respect to $\mathbf{w}$ and setting it equal to zero will result in a cubic term depending on $y(n)$. Through such procedure, it is not possible to isolate the variable $\mathbf{w}$, writing it as a function of the channel. However, a recent result obtained in [15], [16] may simplify such analysis. In these works, the authors show that, under Assumption 2, it is possible to write $J_{\rm CM}$ as

$$J_{\rm CM} = E\{(y(n) - a(n-d))^2(y(n) + a(n-d))^2\} \quad (10)$$

where $a(n)$ is the transmitted symbol and d is a delay. Such manipulation is possible since $E\{a(n-d)^4\} = 1 = R$. Thus, the blind CM criterion may be written as a supervised cost function, depending on the transmitted symbols $a(n-d)$. Using the Cauchy–Schwarz inequality, the authors go even further showing that

$$J_{\rm CM} \leq \sqrt{E\{(y(n) - a(n-d))^4\}E\{(y(n) + a(n-d))^4\}} \quad (11)$$

and thus establishing a superior bound to the CM cost function that depends on fourth-order Wiener like terms. The inequality shown in (11) may be approximated as an equality in regions near a good CM minima, which will lead to good equalization results. Therefore, considering that the equalizer taps are near a minimum solution, where $J_{\rm CM}$ may be considered to assume small values, the following approximation is valid

$$E\{(y(n) - a(n-d))^4\} \cong E\{(y(n) - a(n-d))^2\}^2 \quad (12)$$

which, in other words, consists in saying that the variance of the square error is approximately zero. This is only rigorously true when there is a perfect channel inversion. Note that this is possible when dealing with a decision feedback equalizer structure, even though the result was originally obtained in a linear FIR equalization context [15], [16]. The authors claim that the approximation is valid in conditions near the zero-forcing solution, even though such solution is not attainable in their context. Nevertheless, even when $J_{\rm CM}$ may be considered high, such as around local minima solutions, the authors observe that the result is close to the expected value obtained through simulations. The decision feedback equalizer context is better for validating the considered approximation, since perfect channel inversion is possible. In Section V, we will see that the approximation also remains valid near local minima convergence, enabling the achievement of very accurate channel configurations in which the algorithm will or will not converge to a bad local minimum.

Using (12) in (11) and considering that the second term in the later may be equivalently approximated results in

$$J_{\rm CM} \cong E\{(y(n) - a(n-d))^2\}E\{(y(n) + a(n-d))^2\}. \quad (13)$$

The next step towards the analytical obtention of the optimum solution is to differentiate (13) with respect to $\mathbf{w}$, extending the result obtained in [15] and [16] to the decision feedback equalizer context. First, expanding the first term on the right-hand side of (13), we obtain

$$\begin{aligned} E\{(y(n) - a(n-d))^2\} = ||\mathbf{h}||^2 + \mathbf{w}^T A(\mathcal{P})\mathbf{w} + 1 - 2h_0 \\ + 2\mathbf{w}^T\tilde{T}(\mathcal{P}) - 2\mathbf{h}^T C^T(\mathcal{P})\mathbf{w} \end{aligned} \quad (14)$$

where $A(\mathcal{P})$ and $C(\mathcal{P})$ were defined in (7), $\tilde{T}(\mathcal{P}) = E\{\hat{\boldsymbol{a}}_n a(n-d)\}$ and d was considered equal to zero since this choice leads to a perfect equalization. Differentiating (14) with respect to $\mathbf{w}$ results in

$$\frac{\partial E\{(y(n) - a(n-d))^2\}}{\partial \mathbf{w}} = 2A(\mathcal{P})\mathbf{w} - 2C(\mathcal{P})\mathbf{h} + 2\tilde{T}(\mathcal{P}). \quad (15)$$

We may now turn to (13) and differentiate $J_{\rm CM}$ finding

$$\begin{aligned} \frac{\partial J_{\rm CM}}{\partial \mathbf{w}} \cong\ & A\mathbf{w}(\mathbf{w}^T A\mathbf{w}) - C\mathbf{h}(\mathbf{w}^T A\mathbf{w}) \\ & + 2(C\mathbf{h} - A\mathbf{w})(\mathbf{h}^T C^T\mathbf{w}) + (||\mathbf{h}||^2 + 1)A\mathbf{w} \\ & - 2\tilde{T}(\mathbf{w}^T\tilde{T}) - (||\mathbf{h}||^2 + 1)(C\mathbf{h}) + 2h_0\tilde{T} \end{aligned} \quad (16)$$

where we dropped the dependence with $\mathcal{P}$ to simplify the notation.

The optimal solution $\mathbf{w}_{\mathrm{CM}}^*$ will be given by making (16) equal to zero. This will lead to a system of equations where the number of variables is equal to the number of equations, but each one of these equations will be nonlinear. In order to render the problem more treatable, let us consider the case of three-tap channels, which, under Assumption 3, leads to a feedback filter with two taps. In this case, the two equations from (16) will have the general form

$$\boldsymbol{\alpha}_1 w_1^3 + \boldsymbol{\alpha}_2 w_1^2 + \boldsymbol{\alpha}_3 w_1 + \boldsymbol{\alpha}_4 w_1^2 w_2 + \boldsymbol{\alpha}_5 w_1 w_2 + \boldsymbol{\alpha}_6 w_1 w_2^2 + \boldsymbol{\alpha}_7 w_2 + \boldsymbol{\alpha}_8 w_2^2 + \boldsymbol{\alpha}_9 w_2^3 + \boldsymbol{\alpha}_{10} = \mathbf{0} \quad (17)$$

where the constant vectors $\boldsymbol{\alpha}_i$ depend on $\mathbf{h}$, $A(\mathcal{P})$, $C(\mathcal{P})$, and $\tilde{T}(\mathcal{P})$. This system of equations does not have a closed form solution but, once $\boldsymbol{\alpha}_i$ with $i = 1, \ldots, 10$ are known, it can be solved numerically.

V. Error Surface Analysis: Three Tap Channels

In [14], the authors study the gross convergence properties of a decision directed algorithm in a decision feedback equalizer structure, classifying the possible local minima into two types: the ones resulting from the decision feedback equalizer structure, named delay-type minima, and the ones resulting from the adaptive algorithm itself. Since the first type is carefully studied in [14], we will concentrate on the second type of minima. For DD-DFE, such analysis was well detailed in [13]. Here, we will extend it to the constant modulus algorithm in a decision feedback equalizer structure.

It is well known that the decision device nonlinearity, present in the decision feedback equalizer structure, divides the equalizer parameter space into polytopes, denoted by $\mathcal{P}$ [13]. Considering Assumptions 1–3, the optimum decision device in the sense of the maximum-likelihood criterion is given by (2). The polytopes are limited by hyperplanes that can be defined as

$$\left\{\mathbf{w}_n \in \mathbb{R}^N : \boldsymbol{a}_n^T \mathbf{h} - \mathbf{w}_n^T \hat{\boldsymbol{a}}_n = 0\right\} \quad (18)$$

where $\boldsymbol{a}_n$ and $\hat{\boldsymbol{a}}_n$ vary through all possible combinations of the alphabet symbols according to Assumption 2.

In addition, let us define a state of past symbols and estimates

$$S_n = (a_{n-1}, \ldots, a_{n-N}; \hat{a}_{n-1}, \ldots, \hat{a}_{n-N}). \quad (19)$$

In each polytope, S_n defines a finite state Markov process, with transitions given by the input symbols $a(n)$ [13]. It is interesting to note that inside a polytope $\mathcal{P}$, the Markov chain will be the same for any decision feedback equalizer setting. Considering the step size of the decision feedback equalizer algorithm to be sufficiently small so that the Markov chain will attain its steady state distribution before the algorithm moves significantly around the polytope, we may consider the statistics involved in the adaptation process to be constant within a polytope. Thus, such statistics may be said to be *piecewise constant*.

The steady state statistics of the Markov chain will determine the shape of the cost function inside each polytope, ruling the behavior of the adaptive decision feedback equalizer. Firstly, since blind algorithms are known to perform well in opened-eye situations, let us restrain our attention to closed-eye channels satisfying Assumption 1. Mathematically, if we consider three tap channels, such a condition may be stated as

$$\mathcal{C} = \{(h_1, h_2) \in \mathbb{R}^2 : |h_1| + |h_2| > 1, |h_1| < 1, |h_2| < 1\}. \quad (20)$$

If the adaptive algorithm is initialized at a polytope that presents a local minima, and considering that the algorithms being studied are of the steepest descent type, convergence to that bad solution will be certain. Thus, initialization is crucial to attain a good solution. However, if no *a priori* information on the channel is known, a sensible choice would be to initialize the feedback filter taps with zero. Therefore, a local minima convergence will occur if the center polytope, i.e., the polytope which includes the vector $\mathbf{w} = [0\ 0]^T$, presents a local minima. The center polytope, also considering the case of a three tap channel, may be defined as

$$\mathcal{P}_0 = \{(w_1, w_2) \in \mathbb{R}^2 : |w_1| + |w_2| < |h_1| + |h_2| - 1\}. \quad (21)$$

In [9] and [13], the authors show that, if a minimum is present in the center polytope, it will certainly be a bad local solution. That can be verified by noting that the optimum solution would be given by $\mathbf{w} = [h_1\ h_2]^T$ and such a condition will not satisfy (21).

In this work we are interested in determining a class of channels for which CMA-DFE, under Assumptions 1–3, when initialized at the origin, will present an ill-convergence. Such a result will be compared to that obtained in [9] and [13] for DD-DFE. To achieve such a result, we first need to obtain $A(\mathcal{P})$, $C(\mathcal{P})$, and $\tilde{T}(\mathcal{P})$, in order to determine the system of equations defined in (17). These statistics depend on the channel being considered. The closed-eye channel class $\mathcal{C}$, defined in (20) may be subdivided into four regions:

$$\mathcal{C}_1 = \{\mathbf{h} \in \mathcal{C} : h_1 > 0, h_2 > 0\} \quad (22)$$

$$\mathcal{C}_2 = \{\mathbf{h} \in \mathcal{C} : h_1 < 0, h_2 > 0\} \quad (23)$$

$$\mathcal{C}_3 = \{\mathbf{h} \in \mathcal{C} : h_1 < 0, h_2 < 0\} \quad (24)$$

$$\mathcal{C}_4 = \{\mathbf{h} \in \mathcal{C} : h_1 > 0, h_2 < 0\}. \quad (25)$$

In each of these subregions, [9] and [13] show that the finite state Markov process formed by S_n defined in (19), presents transition probabilities that are invariant to $\mathbf{w} \in \mathcal{P}_0$. This means that the statistics involved in the decision feedback equalizer adaptation at $\mathcal{P}_0$, i.e., $A(\mathcal{P}_0)$, $C(\mathcal{P}_0)$, $T(\mathcal{P}_0)$, and $\tilde{T}(\mathcal{P}_0)$, may be found through the finite state Markov process that depends only on the subregion $\mathcal{C}_i$ that the current channel belongs to and not on the channel itself.

Finally, for any $\mathbf{h} \in \mathcal{C}_i$, with $i = 1, 2, 3, 4$, convergence to local minima after a zero initialization will occur if $\mathbf{w}^*(\mathcal{P}_0) \in \mathcal{P}_0$, where $\mathbf{w}^*(\mathcal{P}_0) = [w_1^*\ w_2^*]^T$ is the solution of (16) using the statistics of $\mathcal{P}_0$. In other words, after (21), convergence to a local minima will occur if

$$|w_1^*| + |w_2^*| < |h_1| + |h_2| - 1. \quad (26)$$

It is important to note that the analysis developed in [13] and reviewed in this section depends only on the definition of the nonlinearity of the decision device. Thus, the results obtained are valid for any adaptation algorithm. In the investigation of which channels satisfy (26) leading to an ill-convergence, the differences in the existence of local minima in $\mathcal{P}_0$ will depend on how $\mathbf{w}^*(\mathcal{P}_0)$ is obtained, which, in its turn, depends on the criterion and algorithm being considered.

A. Statistics Definition in $\mathcal{C}_i$

Having defined subregions $\mathcal{C}_i$ with $i = 1,2,3,4$ in (22) to (25), in order to obtain $\mathbf{w}^*(\mathcal{P}_0)$ for both DD-DFE and CMA-DFE, we need to build the finite state Markov process for each subregion $\mathcal{C}_i$. Its steady state statistics will enable us to obtain the matrices $A(\mathcal{P}_0)$, $C(\mathcal{P}_0)$, $T(\mathcal{P}_0)$, and $\tilde{T}(\mathcal{P}_0)$. The Markov chain states are defined by (19), where a and $\hat{a}$ will vary through all possible combinations of -1 and $+1$. Considering a three tap channel, this will give us a Markov chain with 16 states. Transitions from one state to the other will depend on $a(n)$ and on the nonlinearity defined in (2). It is interesting to note, however, that the statistics are constant for any $\mathbf{w}$ in $\mathcal{P}_0$. This includes the case in which $\mathbf{w} = [0\ 0]^T$. Thus, the transitions may be found by analyzing

$$\hat{a}(n) = \mathrm{sign}(y(n))|_{\mathbf{w}=[0\ 0]^T} = \mathrm{sign}(x(n)) = \mathrm{sign}(\boldsymbol{a}_n^T \mathbf{h}). \quad (27)$$

Through the steady-state Markov chain we are able to obtain the following matrices:

- $\mathcal{C}_1$:

$$A(\mathcal{P}_0) = \begin{bmatrix} 1 & 0.5 \\ 0.5 & 1 \end{bmatrix},$$
$$C(\mathcal{P}_0) = \begin{bmatrix} 0 & 0.5 & 0.5 \\ 0 & 0 & 0.5 \end{bmatrix},\ T(\mathcal{P}_0) = \begin{bmatrix} 0.5 \\ 0.25 \end{bmatrix} \quad (28)$$

- $\mathcal{C}_2$:

$$A(\mathcal{P}_0) = \begin{bmatrix} 1 & -0.5 \\ -0.5 & 1 \end{bmatrix},$$
$$C(\mathcal{P}_0) = \begin{bmatrix} 0 & 0.5 & -0.5 \\ 0 & 0 & 0.5 \end{bmatrix},\ T(\mathcal{P}_0) = \begin{bmatrix} -0.5 \\ 0.25 \end{bmatrix} \quad (29)$$

- $\mathcal{C}_3$:

$$A(\mathcal{P}_0) = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix},$$
$$C(\mathcal{P}_0) = \begin{bmatrix} 0 & 0.5 & -0.5 \\ 0 & 0 & 0.5 \end{bmatrix},\ T(\mathcal{P}_0) = \begin{bmatrix} 0 \\ -0.25 \end{bmatrix} \quad (30)$$

- $\mathcal{C}_4$:

$$A(\mathcal{P}_0) = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix},$$
$$C(\mathcal{P}_0) = \begin{bmatrix} 0 & 0.5 & 0.5 \\ 0 & 0 & 0.5 \end{bmatrix},\ T(\mathcal{P}_0) = \begin{bmatrix} 0 \\ -0.25 \end{bmatrix}. \quad (31)$$

For every subregion $\mathcal{C}_i$, $\tilde{T}(\mathcal{P}_0)$ will be equal to zero what was already expected under Assumption 2.

B. Defining the Class of Bad Channels for CMA-DFE

Using the matrices defined in Section V-A, we must now obtain the optimum CM solutions through the system of equations given by (17) and test if they verify condition (26). Since there is no closed form solution for $\mathbf{w}^*_{\mathrm{CM}}$ through (17), we will solve the system of equations numerically. The dashed area in Fig. 2

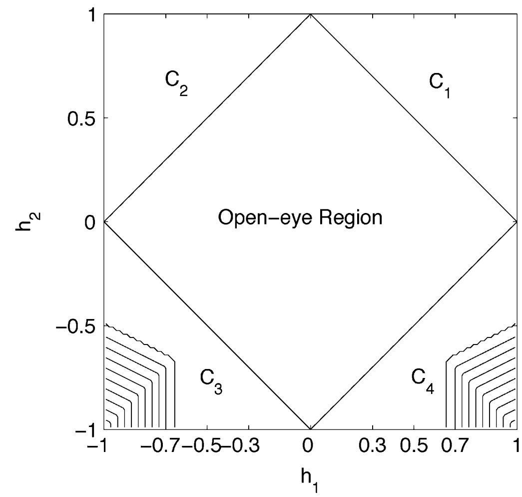

Fig. 2. Class of bad channels for CMA-DFE.

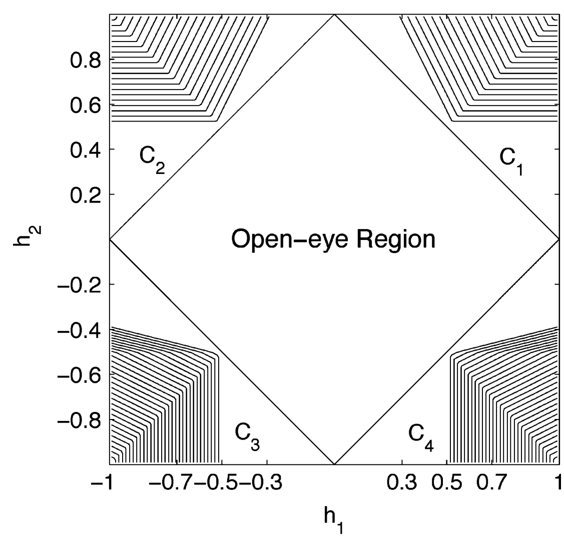

Fig. 3. Class of bad channels for DD-DFE.

shows the regions in which there exists a solution $\mathbf{w}_{\mathrm{CM}}$ satisfying (26), that is, in which convergence to a local minimum will occur. Note that solving (17) gives us several solutions for $\mathbf{w}_{\mathrm{CM}}$. Thus, the ill-convergence will occur if at least one of the given solutions satisfies (26). The figure is given in terms of h_1 and h_2, under Assumptions 1–3 and with $\mathbf{h} \in \mathcal{C}$. Channels belonging to the dashed area are named *bad channels* since they lead to the ill-convergence of CMA-DFE.

In the sequence, we also recover the results obtained for the DD-DFE in [9] and [13]. In this case, $\mathbf{w}^*_{\mathrm{DD}}$ is obtained through (7) with the same matrices given in Section V–A. Fig. 3 illustrates the cases where $\mathbf{w}^*_{\mathrm{DD}} \in \mathcal{P}_0$.

Comparing with the CMA-DFE, given by Fig. 2, we may confirm the superior robustness of the CMA with respect to the DD algorithm. Such a difference in performance was briefly observed in [9], without major analysis. We have now numerically verified that the class of channels leading to the ill-convergence of the CMA-DFE is considerably smaller than that of the DD-DFE. Mainly, for channels belonging to $\mathcal{C}_1$ and $\mathcal{C}_2$,

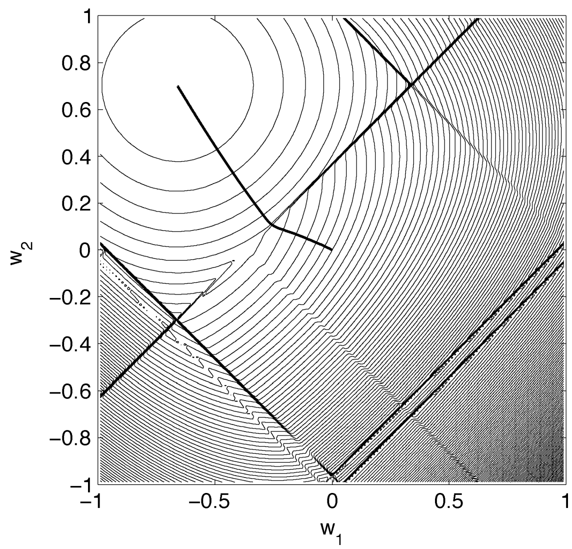

Fig. 4. CMA-DFE error surface and convergence for $\mathbf{h}_1 = [1\ -0.66\ 0.7] \in \mathcal{C}_2$, $\mu = 1 \times 10^{-4}$.

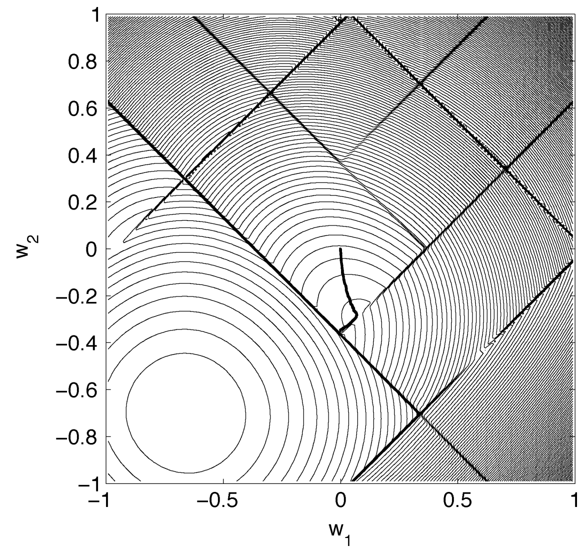

Fig. 5. CMA-DFE error surface and convergence for $\mathbf{h}_2 = [1\ -0.66\ -0.7] \in \mathcal{C}_3$, $\mu = 1 \times 10^{-4}$.

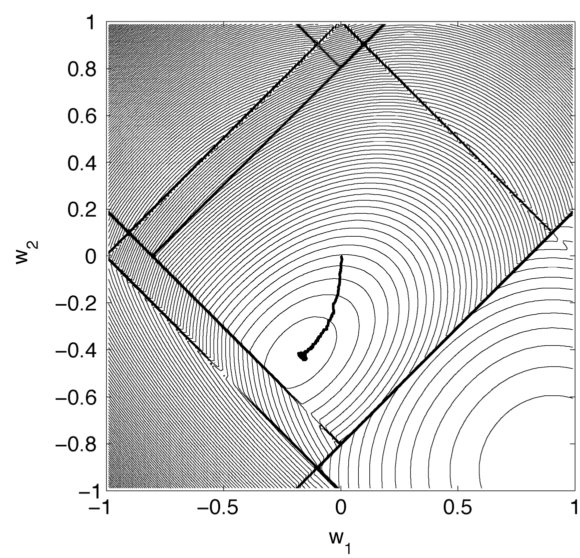

Fig. 6. CMA-DFE error surface and convergence for $\mathbf{h}_3 = [1\ 0.9\ -0.9] \in \mathcal{C}_4$, $\mu = 1 \times 10^{-4}$.

CMA-DFE will not present a local minimum in $\mathcal{P}_0$. Thus, the algorithm will certainly move towards another polytope, which may contain a global solution. As for channels belonging to $\mathcal{C}_3$ and $\mathcal{C}_4$, we may observe that the class of bad channels for CMA-DFE is still smaller than that presented by DD-DFE.

It is also interesting to note that regions $\mathcal{C}_1$ and $\mathcal{C}_2$, in which CMA-DFE does not present a local minima in $\mathcal{P}_0$, are given by minimum phase channels. On the other hand, regions $\mathcal{C}_3$ and $\mathcal{C}_4$, in which CMA-DFE presents local minima in $\mathcal{P}_0$, are given by non-minimum phase channels, i.e., such channels present one root at the exterior of the unit circle.

It should be emphasized that the results obtained above consider a decision feedback equalizer implemented using only a feedback filter. Independently of the channel, when considering a feedforward filter well projected in terms of length and initialization, the resultant channel viewed by the feedback filter will probably be a minimum phase channel. In such a situation, as shown above, the CMA-DFE will converge to the optimum solution or to a delay-type minima, which is sufficient to open the eye. However, the choice of such parameters may engender some difficulties since usually the channel is not known. It should also be noted that a theoretical analysis of such behavior is still very difficult.

C. Simulation Example

Let us consider the channels $\mathbf{h}_1 = [1\ -0.66\ 0.7]$ and $\mathbf{h}_2 = [1\ -0.66\ -0.7]$, where $\mathbf{h}_1$ belongs to $\mathcal{C}_2$ and $\mathbf{h}_2$ belongs to $\mathcal{C}_3$. Figs. 4 and 5 show the CMA-DFE error surface and convergence respectively for $\mathbf{h}_1$ and $\mathbf{h}_2$. The step size used was equal to 1×10^{-4}. As expected, the error surface in Fig. 4 does not present a local minimum in $\mathcal{P}_0$ and the algorithm converges to the optimum solution perfectly inverting the channel. On the other hand, channel $\mathbf{h}_2$ is a bad channel near the border of the dashed area shown in Fig. 2. For this reason, Fig. 5 shows a shallow local minimum in $\mathcal{P}_0$. Channels closer to the corner present deeper local minima as can be seen by Fig. 6, where the channel was considered to be $\mathbf{h}_3 = [1\ 0.9\ -0.9]$. Anyway, this shallow minimum is sufficient to hold the convergence of the algorithm when small step sizes are used [21].

Fig. 7 shows the result obtained for DD-DFE in channel $\mathbf{h}_1$. As expected, the error surface presents a local minimum in $\mathcal{P}_0$ confirming the better performance of the CMA-DFE in this case.

Considering channels in $\mathcal{C}_3$ that are not bad channels but are also close to the border of the shaded region in Fig. 2, we may observe a transition interval for which the error surface will not present a local minimum in $\mathcal{P}_0$ but it has a rather flat form near the edges that define the center polytope. An example of such a channel is $\mathbf{h}_4 = [1\ -0.64\ -0.7]$. In this case, if we consider the step size μ smaller than 10^{-5}, the algorithm is not able to cross to another polytope and converge to the global minimum. It stays trapped in the border of $\mathcal{P}_0$. This situation can be seen in Fig. 8. Note, however, that a very small step size must be used to achieve such a behavior. For $\mu > 1 \times 10^{-5}$, the algorithm escapes from $\mathcal{P}_0$ and converges to the global minimum. As we

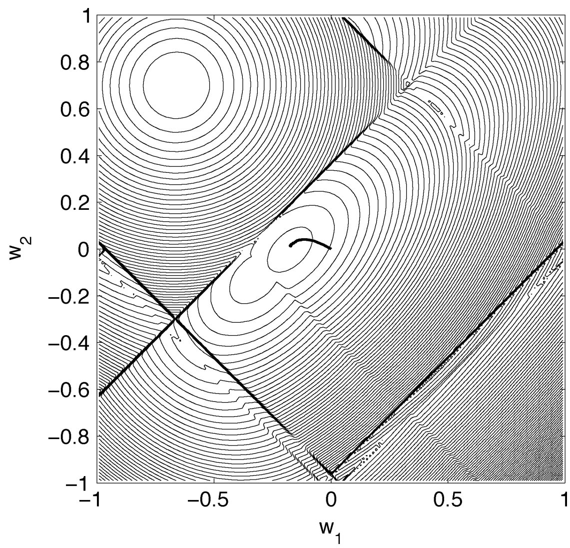

Fig. 7. DD-DFE error surface and convergence for $\mathbf{h}_1 = [1\ -0.66\ 0.7] \in \mathcal{C}_2$, $\mu = 1 \times 10^{-4}$.

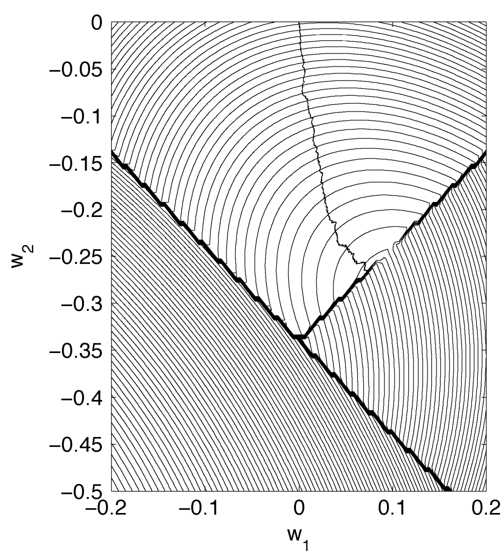

Fig. 8. CMA-DFE error surface and convergence for $\mathbf{h}_4 = [1 - 0.64 - 0.7]$, $\mu = 1 \times 10^{-5}$.

continue moving away from the shaded areas in Fig. 2, the minimum in $\mathcal{P}_0$ vanishes completely and the algorithm will always move to another polytope that may present a local or global solution. The analysis above may easily be extended to channels belonging to $\mathcal{C}_4$, leading to the same results stated above.

It is also important to note that such results demonstrate that the approximation given by (12) and used to obtain the system of equations shown in (16), provides excellent accuracy in the sense of defining the classes of bad channels.

VI. Generalization

Generalizing the number of channel coefficients, N, it is possible to achieve similar results as the ones developed for three tap channels. Once again, we will restrain our attention to closed-eye channels, since decision directed and constant modulus algorithms converge well in opened-eye situations assuming a center initialization. It was seen, for $N = 2$, that the Markov chain ruling the statistics in the center polytope $\mathcal{P}_0$ for a closed-eye channel depended only on the class $\mathcal{C}_i$ the channel belonged to. Once the class is defined, the statistics are constant over $\mathcal{P}_0$. This is no longer true for higher values of N. For this reason, in [13], the authors consider a more restrictive class, the *corner*-class of channels given by

$$\mathcal{C}^{(N)} = \left\{ (h_1, \ldots, h_N)^T \in \mathbb{R}^N : 1 - \frac{1}{N} < |h_i| < 1 \right\} \tag{32}$$

which satisfies the property that for any $\mathbf{h} \in \mathcal{C}^{(N)}$, $\mathbf{w}^*$ is not in $\mathcal{P}_0$ and that the Markov chain statistics will only depend on which corner the channel parameters belong to. Thus, once again, the statistics associated with $\mathcal{P}_0$ will be constant.

It can be easily shown that if a channel is in $\mathcal{C}^{(N)}$, the desired global solution can not be in $\mathcal{P}_0$ [13]. Therefore, if the central polytope presents a solution, it will certainly be a local minimum. Each corner in the channel parameter space may be denoted as $\mathcal{C}_i^{(N)}$ where

$$i = \frac{1}{2}(2^N - 1 + 2^{N-1}\lambda_1 + 2^{N-2}\lambda_2 + \cdots + \lambda_N) \tag{33}$$

with $\lambda_i = \mathrm{sign}(h_i)$. Thus, knowing which corner-class $\mathcal{C}_i^{(N)}$ the channel belongs to, the statistics of the center polytope may be obtained and the optimum feedback filter taps w_i^* may be calculated using (7) for DD-DFE and through the solution of (17) for the CMA-DFE. Ill-convergence will occur if the following condition is satisfied [13]:

$$\sum_{i=1}^{N} |w_i^*| - \min_{\nu_j} \left| \sum_{j=0}^{N} \nu_j h_j \right| < 0 \tag{34}$$

where $\nu_j = a_{n-j}$, belonging to $\{-1, +1\}$. More details on this analysis may be found in [13].

It is important to note that, even though the result given by (34) closely follows what was done for the three-tap channel case, when considering CMA-DFE, the obtention of $\mathbf{w}_{\mathrm{CM}}^*$ may quickly become complex as N grows. The system given by (17) contains nonlinear third-order equations on the feedback filter coefficients, including crossed terms, that may not be solved through a closed form expression.

VII. Conclusion

In this work, we have extended the class of channels for which a decision feedback equalizer, implemented using only a feedback filter, will converge to a local minimum when initialized at the origin of the feedback filter parameter space. Such a result had already been developed for the DD-DFE. Extending the idea to the CMA-DFE, we have shown that, for three-tap closed-eye channels, the class that will lead to ill-convergence is much smaller than that presented by the DD-DFE, confirming the greater robustness of the former. In particular, the CMA-DFE does not present local minima for minimum phase channels. In previous papers, such a result had been observed only through simulations without being properly explained, specially due to the analytical difficulty in obtaining the optimum filter tap solution as a function of a given channel for the CM criterion. To achieve such solution, we use a recently developed approximation first proposed in a linear FIR context, extending it to the decision feedback equalizer scenario. Generalization to a channel of order $N+1$ is briefly discussed, since, in this case, finding the optimum solution for the filter taps in a CMA-DFE may become quite complex.

The absence of a local minimum that results in the ill-convergence of a null initialized feedback filter for minimum phase channels constrained to Assumptions 1–3 implies an interesting characteristic: it possibly guarantees the good operation of a CMA-DFE when using a sufficiently large feedforward filter with an adequate initialization. We may argue in this sense since the constant modulus algorithm will lead to a solution similar to the mean square error criterion, and the feedforward filter may behave similarly to the whitened matched filter. A whitened matched filter will turn the non minimum phase channel, viewed by the feedback filter, into a minimum phase channel. Therefore, assuming a spike initialization of the feedforward filter, the feedback filter starts its adaptation observing a non-minimum phase channel, possibly resulting in an initial ill-convergence but, with the adaptation of the feedforward filter, the resultant channel observed by the feedback filter should become closer to a minimum phase channel avoiding the ill-convergence of the decision feedback equalizer. Even if there is a residual precursor response, this should not be a problem, since it should be small and can be approximated as Gaussian noise. Noise also helps the convergence of the equalizer. Nonetheless, even if a sufficiently large feedforward filter is used with a reasonable initialization, a caveat must be made regarding the convergence of such CMA-DFE: We have no guarantee that the feedforward filter will indeed approximate a whitened matched filter. An example of such a case is the equalizer convergence to a degenerated solution, which fortunately can be avoided by the use of constraints [11], [22]. In addition, even if the feedforward filter approximates a whitened matched filter, the equalizer may converge to a delay-type minima [14].

REFERENCES

[1] J. G. Proakis and M. Salehi, *Digital Communications*, 5th ed. New York: McGraw-Hill, 2008.

[2] G. Forney, Jr., "Maximum-likelihood sequence estimation of digital sequences in the presence of intersymbol interference," *IEEE Trans. Inf. Theory*, vol. 18, no. 3, pp. 363–378, May 1972.

[3] B. Sklar, "How I learned to love the trellis," *IEEE Signal Process. Mag.*, vol. 20, no. 3, pp. 87–102, May 2003.

[4] G. Ungerboeck, "Adaptive maximum-likelihood receiver for carrier-modulated data-transmission systems," *IEEE Trans. Commun.*, vol. 22, no. 5, pp. 624–636, May 1974.

[5] M. E. Austin, "Decision-feedback equalization for digital communication over dispersive channels," MIT Lincoln Lab. Research Lab. Tech. Rep. No. 467, 1967.

[6] C. Belfiore and J. H. Park, Jr., "Decision feedback equalization," *Proc. IEEE*, vol. 67, no. 8, pp. 1143–1156, Aug. 1979.

[7] D. Godard, "Self-recovering equalization and carrier tracking in two-dimensional data communication systems," *IEEE Trans. Commun.*, vol. COM-28, no. 11, pp. 1867–1875, 1980.

[8] J. Treichler and B. Agee, "New approach to multipath correction of constant modulus signals," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. ASSP-31, no. 2, pp. 459–472, 1983.

[9] R. Casas, Z. Ding, R. Kennedy, C. R. Johnson, Jr., and R. Malamut, "Blind adaptation of decision feedback equalizers based on the constant modulus algorithm," in *Proc. 29th Asilomar Conf. Signals, Systems, Computers*, 1996, vol. 1, pp. 698–702.

[10] T. Endres, C. Strolle, S. Hulyalkar, T. Chaffer, A. Shah, M. Gittings, C. Hollowell, A. Bhaskaran, J. Roletter, and B. Paratore, "Carrier independent blind initialization of a DFE," in *Proc. 2nd IEEE Signal Processing Workshop Signal Processing Advances in Wireless Communications*, 1999, pp. 239–242.

[11] C. Papadias and A. Paulraj, "Decision-feedback equalization and identification of linear channels using blind algorithms of the bussgang type," in *Proc. 29th Asilomar Conf.*, 1996, pp. 335–340.

[12] R. Casas, C. R. Johnson, Jr., J. Harp, and S. Caffee, "On initialization strategies for blind adaptive DFE," in *Proc. IEEE Wireless Communication Networking Conf.*, 1999, vol. 2, pp. 792–796.

[13] R. Casas, C. R. Johnson, Jr., R. Kennedy, Z. Ding, and R. Malamut, "Blind adaptive decision feedback equalization: A class of channels resulting in ill-convergence from a zero initialization," *Int. J. Adaptive Control Signal Process.*, vol. 12, pp. 173–193, 1998.

[14] R. Kennedy, B. Anderson, and R. Bitmead, "Blind adaptation of decision feedback equalizers: Gross convergence properties," *Int. J. Adapt. Control Signal Process.*, vol. 7, pp. 497–523, 1993.

[15] R. Suyama, R. R. Attux, J. M. Romano, and M. Bellanger, "On the relationship between least squares and constant modulus criteria for adaptive filtering," in *Proc. 37th Asilomar Conf. Signals, Systems, Computers*, Pacific Grove, 2003, vol. 2, pp. 1293–1297.

[16] R. Suyama, R. R. Attux, J. M. Romano, and M. Bellanger, "Relations entre les critères du module constant et de Wiener," presented at the 19th GRETSI Symp. Signal Image Processing, Paris, France, Sep. 8–11, 2003.

[17] P. Regalia and M. Mboup, "Properties of some blind equalization criteria in noisy multi-user environments," *IEEE Trans. Signal Process.*, vol. 49, no. 12, pp. 3112–3122, 2001.

[18] P. Schniter and L. Tong, "Bounds for the MSE performance of constant modulus estimators," *IEEE Trans. Inf. Theory*, vol. 46, no. 7, pp. 2544–2560, 2000.

[19] H. Zeng, L. Tong, and C. R. Johnson, Jr., "Relationships between the constant modulus and Wiener receivers," *IEEE Trans. Inf. Theory*, vol. 44, no. 4, pp. 1523–1538, 1998.

[20] H. Zeng, L. Tong, and C. R. Johnson, Jr., "An analysis of constant modulus receivers," *IEEE Trans. Signal Process.*, vol. 47, no. 11, pp. 2990–2999, 1999.

[21] J. E. Mazo, "Analysis of decision-directed equalizer convergence," *Bell Syst. Tech. J.*, vol. 59, pp. 1857–1876, 1980.

[22] L. Szczecinski and A. Gei, "Blind decision feedback equalisers, how to avoid degenerative solutions," *Signal Process.*, vol. 82, pp. 1675–1693, 2002.

Aline Neves received the B.S. and M.S. degree in electrical engineering from the State University of Campinas (UNICAMP), Brazil, in 1999 and 2001, respectively, and the Ph.D. degree, also in electrical engineering, from the University René Descartes (Paris V), Paris, France, in 2005.

She is currently an Assistant Professor at the Engineering, Modeling and Applied Social Science Center of the Federal University of ABC (CECS/UFABC), Santo André, Brazil. Her research interests include equalization, channel estimation, source separation, and information theoretic learning.

Cristiano Panazio (M'05) received the B.Sc. and M.Sc. degrees in electrical engineering from the State University of Campinas (UNICAMP), Brazil, in 1999 and 2001, respectively, and the Ph.D. degree, also in electrical engineering, from the Conservatoire National des Arts Métiers (CNAM), Paris, France, in 2005.

In 2006, he became Assistant Professor at the Escola Politécnica of the University of São Paulo (EPUSP), São Paulo, Brazil. His research interests include equalization, multicarrier modulations, and spread spectrum techniques.