Analysis of Blind Decision Feedback Equalizer Convergence: Interest of a Soft Decision

S. Cherif, S. Marcos, M. Jaidane

Abstract—In this paper the behavior of the decision feedback equalizers (DFEs) adapted by the decision-directed or the constant modulus blind algorithms is presented. An analysis of the error surface of the corresponding criterion cost functions is first developed. With the intention of avoiding the ill-convergence of the algorithm, the paper proposes to modify the shape of the cost function error surface by using a soft decision instead of the hard one. This was shown to reduce the influence of false decisions and to smooth the undesirable minima. Modified algorithms using the soft decision during a pseudo-training phase with an automatic switch to the properly tracking phase are then derived. Computer simulations show that these modified algorithms present better ability to avoid local minima than conventional ones.

Keywords—Blind DFEs, decision-directed algorithm, constant modulus algorithm, cost function analysis, convergence analysis, soft decision.

I. INTRODUCTION

IN multi-point transmission systems and in radio communication, the problem of channel equalization is particularly difficult. Indeed, these systems are characterized by a random user access and by a time-varying transmission channel [1], [2]. To satisfy utilization constraints, different classes of blind or self-adaptive equalization techniques, where training phase is avoided, are currently the purpose of many applied researches, so as to define for example WLAN, UWB or WiMAX norms. Among them, the Bussgang-type algorithms are initially introduced and analyzed for a transversal structure of the equalizer [3], [4]. This structure presents the disadvantage of providing limited performances in the case of strongly dispersive or rapidly time-varying channels, for what one prefers the Decision Feedback Equalizers (DFEs) [5], [6], [7].

The functioning of the different blind equalization algorithms is in general disturbed by the existence of local minima, preventing a correct equalization of the transmission channel [8], [9]. For blind equalizers with linear transversal

structure, Ding et al. [10] have emphasized the false convergence to this type of minima. They showed that baud-rate blind equalization algorithms might converge to undesirable stable equilibriums due to different reasons. One is the use of an FIR filter as an equalizer. The others are due to the cost function adopted by the blind algorithm itself, even implemented with double infinite equalizers.

Indeed, to invert a channel filter of Finite Impulse Response (FIR), it would be necessary to use an equalizer filter of Infinite Impulse Response (IIR). In decision feedback equalization, the inversion of the FIR channel filter by a FIR filter equalizer becomes possible. However, despite this perfect modeling of the transmission channel, the error surfaces of the different criteria, corresponding to blind DFE algorithms, also present a non-convex form that can, possibly, present local minima [11].

The purpose of the paper is to outline the interest of the use of soft decision to avoid the ill-convergence of DFEs. The soft decision is often used in neural networks and in communication techniques to reduce the effect of the error propagation. In this paper we give other sense to the use of the soft decision since it is used to modify the error surface. The goal is to attenuate the local minima and then to ensure the convergence to the desired minimum.

The paper is organized as follows. Section 2 contains the system model. We point out, in Section 3, the approach used for the analysis of a DFE in presence of noiseless channel. In Section 4, the error surfaces of the decision-directed and constant modulus algorithm cost functions are investigated. We prove, in Section 5, that the use of a soft decision instead of the decision device during an initialization period, improves the performances of adaptive DFEs, avoiding ill-convergence. Modified decision-directed and constant modulus algorithms achieving an automatic switch between the initialization phase and the properly tracking phase are the proposed. Finally, Computer simulations show, in Section 6, how the use of the soft decision improves the performances of the DD and the constant modulus algorithms.

II. SYSTEM MODELING

We here consider an equivalent discrete time baseband model of the digital transmission channel including the physical channel and the corresponding transmitter and receiver. Fig. 1 illustrates the system under consideration. The transmitted data a_n are supposed independent and identically distributed (iid) taking, with equal probability, a finite number

of levels: $\pm 1, \pm 3, \cdots, \pm(K-1)$; where K is the number of levels. The signal y_n received at the output of the channel is:

$$y_n = a_n + \Theta^T A_n + v_n, \qquad (1)$$

where $\Theta = (\theta_1, \cdots, \theta_{N-1})^T$, $A_n = (a_{n-1}, \cdots, a_{n-N+1})^T$ and v_n an additive white Gaussian noise of zero mean and variance σ_v^2. This equation shows the presence of a direct path in the transmission channel.
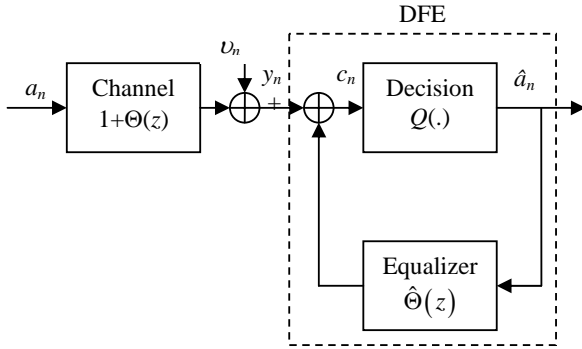

Fig. 1 Equivalent discrete time baseband model of the digital transmission channel including a decision feedback equalizer

At the receiver, the input of the decision device is written:

$$c_n = y_n - \hat{\Theta}^T \hat{A}_n, \qquad (2)$$

where $\hat{\Theta} = (\hat{\theta}_1, \cdots, \hat{\theta}_{M-1})^T$, $\hat{A}_n = (\hat{a}_{n-1}, \cdots, \hat{a}_{n-M+1})^T$, and $\hat{a}_n = Q(c_n)$, with $Q(x) = \sum_{k=-K/2+1}^{K/2-1} \text{sgn}(x+2k)$ ($\text{sgn}(x)=1$ if $x>0$, $\text{sgn}(x)=-1$ if $x<0$ and $\text{sgn}(0)=0$). Transmission attenuation is supposed recovered by the receiver.

A more typical and indeed more general DFE structure usually consists of a feedforward FIR filter followed by a feedback FIR filter. However, in this paper, only the Feedback FIR filter is considered for two reasons. Firstly, in the more general structure it is usually possible to separate the adaptation of the feedforward section from the feedback one. Secondly, the aim of this paper is to highlight the performances of the feedback section when soft decision is used. However in Section 6, for computer simulations, the general form of the DFE with feedforward FIR filter will be considered for performance analysis.

In adaptive equalization, the optimization of the DFE is usually performed during a training phase when a transmitted data sequence has to be known by the receiver. In blind equalization, the first idea was to use $\hat{a}_n$ instead of a_n from the beginning of the adaptation, giving place to the Decision-Directed (DD) criterion [4] whose cost function is given by:

$$J_d = E\{(\hat{a}_n - c_n)^2\}. \qquad (3)$$

The LMS adaptive algorithm that derives from (2) and (3) is the following:

$$\hat{\Theta}_d(n+1) = \hat{\Theta}_d(n) - \mu_d(\hat{a}_n - c_n)\hat{A}_n. \qquad (4)$$

The most famous algorithm for blind equalization is the Constant Modulus Algorithm (CMA), which is a special case of the Godard algorithm [12]. It derives from the following cost function:

$$J_c = \frac{1}{4} E\{(|c_n|^2 - \gamma)^2\}, \quad \gamma = \frac{E\{|a_n|^4\}}{E\{|a_n|^2\}}. \qquad (5)$$

The corresponding algorithm is:

$$\hat{\Theta}_c(n+1) = \hat{\Theta}_c(n) + \mu_c c_n (c_n^2 - \gamma)\hat{A}_n. \qquad (6)$$

In this paper, we limit our study to the decision-directed and the constant modulus algorithms and to the corresponding cost functions. Step sizes μ_d and μ_c are chosen to ensure algorithm convergences.

III. PROBLEM FORMULATION

A. Global Approach of the Error Surface Analysis

In the context of channel equalization in the presence of finite alphabet signals, criteria of blind optimization applied to the DFEs present a multimodal error surface, having several minima: global minima corresponding to the desired functioning of the equalizer and local minima corresponding to undesirable one.

To localize the points of potential convergence of the blind DFE, we draw one's inspiration from the approach proposed in [11] and [13]. In these references, the authors establish the relationship between regions in parameter space $D_{\hat{\Theta}}$, called polytopes, and finite state Markov processes which completely specify the statistics which affect the location of convergence points of the blind algorithm. A partition of the space of the parameter vector $\hat{\Theta}$ in a series of polytopes separated by hyperplanes is presented. The objective is to return the global analysis of the algorithm, or the criterion in question, to a local analysis proper to each polytope.

To apply this approach, we are interested first in a noiseless channel and in an equalizer impulse response of length $M = N$. The hyperplanes are then defined by the set:

$$D_{\hat{\Theta}} = \left\{ \hat{\Theta} : c_n = \left(a_n + \Theta^T A_n - \hat{\Theta}^T \hat{A}_n \right) \in S_K \right\}; \qquad (7)$$

with a_n and $\hat{a}_n$ taking values in $\{\pm 1, \pm 3, \cdots, \pm(K-1)\}$ and $S_K = \{(2-K), (4-K), \cdots, (K-2)\}$.

Each choice of a_{n-i} and $\hat{a}_{n-i}$ in $\{\pm 1, \pm 3, \cdots, \pm(K-1)\}$ and of c_n in: $(2-K), (4-K), \cdots, (K-2)$ defines a hyperplane. These hyperplans yield a number of manifolds, referred to as polytopes in [11]. In each polytope, c_n is in the same region of decision (i.e. $Q(c_n)$ takes the same value). Within each polytope, sequences $\{a_n\}$ and $\{\hat{a}_n\}$ are independent of the parameter vector $\hat{\Theta}$.

Thus, the interest of this partition is that each polytope cannot present more than one convergence point. The study of the convergence is then simplified by taking position each time into a given polytope. Finally, note that these results are independent of the blind criterion under study, so they are

valid for all blind criteria especially for the DD and CMA that we are going to analyze in what follows.

To make lucid the presentation of the problematic, we here consider the case of $K = 2$ ($a_n = \pm 1$), and a noiseless channel and an equalizer of order $N = M = 1$ (we note $\theta = \theta_1$ and $\hat{\theta} = \hat{\theta}_1$).

Thereafter, we consider two particular running cases of the equalizer. The first concerns the desired behavior of the equalizer, in other words the case where the decisions are correct:

$$E\{\hat{a}_n a_n\} = 1; \quad \text{or} \quad \hat{a}_n = a_n, \quad \forall n > N_1 \quad (N_1 \in \text{IN}). \quad (8)$$

The second concerns the undesirable behavior, corresponding to the case where the decisions $\hat{a}_n = \text{sgn}(c_n)$ are independent of the values of a_n, $\forall n > N_2$ ($N_2 \in \text{IN}$), that is:

$$E\{\hat{a}_n a_n\} = 0. \quad (9)$$

By taking into account the above mentioned conditions, the space of parameters $D_{\hat{\Theta}}$, defined by (7), is reduced to an axis of parameter $\hat{\theta}$:

$$D_{\hat{\Theta}} = \left\{ \hat{\theta} : a_n + \theta a_{n-1} - \hat{\theta} \hat{a}_{n-1} = 0 \right\}. \quad (10)$$

where $a_n, a_{n-1}, \hat{a}_{n-1} \in \{-1, 1\}$.

Thus, hyperplanes are reduced to four vertical lines: $\hat{\theta} = \pm 1 \pm \theta$, which separate five polytopes. We show in the following that among these polytopes some correspond to correct decision regions and others correspond to incorrect decision regions.

In polytopes corresponding to a correct decision, we have the equality $\hat{a}_n = \text{sgn}(c_n) = a_n$. Thus, the following condition, according to (1) and (2), must be verified:

$$| \theta a_{n-1} - \hat{\theta} \hat{a}_{n-1} | < | a_n |, \quad \forall a_n, \forall a_{n-1} \quad \text{with} \quad \hat{a}_{n-1} = a_{n-1}. \quad (11)$$

As $| a_n | = 1$, expression (11), that defines the polytopes of correct decisions, becomes:

$$-1 + \theta < \hat{\theta} < 1 + \theta. \quad (12)$$

By analogy with the above analysis for correct decision polytopes, we deduce the expressions:

$$| \hat{\theta} | > 1 + | \theta |,$$
$$| \hat{\theta} | < -1 + | \theta | \quad \text{or} \quad | \hat{\theta} | > 1 + | \theta |; \quad (13)$$

which define polytopes of incorrect decisions ($E\{\hat{a}_n a_n\} = 0$), for minimum phase and maximum phase channels respectively.

These two regions of functioning do not cover the totality of the error surface. There are other regions where the decision is uncertain. Although these regions correspond globally to false decision regions, we cannot assert that inside them $E\{\hat{a}_n a_n\} = 0$.

B. Error Surface Analysis

The expression of the cost functions corresponding to the DD and CMA criteria, using (1), (2), (3) and (5), are respectively:

$$
\begin{aligned}
J_d &= 2 + \theta^2 + \hat{\theta}^2 - 2\theta\hat{\theta} E\{\hat{a}_{n-1} a_{n-1}\} - 2E\{\hat{a}_n a_n\} \\
&\quad - 2\theta E\{\hat{a}_n a_{n-1}\} + 2\hat{\theta} E\{\hat{a}_n \hat{a}_{n-1}\};
\end{aligned}
$$

$$
\begin{aligned}
J_c &= \left(\theta\hat{\theta}\right)^2 + \left(\theta^2 + \hat{\theta}^2\right)\left(1 + \frac{1}{4}\theta^2 + \frac{1}{4}\hat{\theta}^2\right) \\
&\quad - \theta\hat{\theta}\left(2 + \theta^2 + \hat{\theta}^2\right) E\{\hat{a}_{n-1} a_{n-1}\} \\
&\quad - \hat{\theta}\left(3\theta^2 + \hat{\theta}^2\right) E\{a_n \hat{a}_{n-1}\} \\
&\quad + \theta\left(\theta^2 + 3\hat{\theta}^2\right) E\{a_n a_{n-1}\}.
\end{aligned}
\quad (14)
$$

The calculation of the terms $E\{.\}$ is made taking into accounts the two cases of correct and incorrect decisions separately. In the case of correct decisions, the terms $E\{a_n a_{n-1}\}$, $E\{\hat{a}_n a_{n-1}\}$, $E\{a_n \hat{a}_{n-1}\}$ and $E\{\hat{a}_n \hat{a}_{n-1}\}$ are null since we have $\hat{a}_n = a_n$, $\hat{a}_{n-1} = a_{n-1}$ and that the sequence $\{a_n\}$ is iid ($E\{a_n a_{n-i}\} = 0$, if $i \neq 0$). Also, the values of the remaining terms are $E\{\hat{a}_n a_n\} = E\{\hat{a}_{n-1} a_{n-1}\} = 1$. Thus, the expressions of the cost functions are simplified to:

$$
\begin{aligned}
J_d^G &= \left(\theta - \hat{\theta}\right)^2, \\
J_c^G &= \frac{1}{4}\left(\theta - \hat{\theta}\right)^2 \left[\left(\theta - \hat{\theta}\right)^2 + 4\right];
\end{aligned}
\quad (15)
$$

where the index G makes reference to the global minimum that is included potentially in the polytope defined by (12), corresponding to correct decisions.

In the case of incorrect decisions, the terms $E\{\hat{a}_n a_n\}$ and $E\{\hat{a}_{n-1} a_{n-1}\}$ are null. The term $E\{a_n a_{n-1}\}$ is also null since the data $\{a_n\}$ are iid; whereas the other terms in (14) take different values according to the considered polytope among the polytopes defined by (13), corresponding to incorrect decisions. The calculus of these different values is given in appendix A and appendix B. Consequently, the expressions of the cost functions also depend on the considered polytopes. We thus obtain:

$$
J_d^L = \begin{cases}
1 + \theta^2 + \left(1 + \hat{\theta}\right)^2, & \text{if} \quad \hat{\theta} < -1 - | \theta |, \\
1 + \left(1 - | \theta |\right)^2 + \hat{\theta}^2, & \text{if} \quad | \hat{\theta} | < -1 + | \theta |, \\
1 + \theta^2 + \left(1 - \hat{\theta}\right)^2, & \text{if} \quad 1 + | \theta | < \hat{\theta};
\end{cases}
\quad (16)
$$

$$
J_c^L = \frac{1}{4}\left(\theta^2 + \hat{\theta}^2\right)\left(\theta^2 + \hat{\theta}^2 + 4\right) + \theta^2 \hat{\theta}^2.
$$

The index L is referred to the local minimum that may be included in polytopes corresponding to incorrect decisions. Note that the second polytope of (16), corresponding to the DD criterion, does not exist in the case of minimum phase channels (equation (13)).

C. Minima of the Cost Functions

Within each polytope, the DD and CMA criteria present possible global or local minima according to whether it concerns correct or faulty decisions. From (15), we derive expressions of the global minima: $\hat{\theta}_d^G = \hat{\theta}_c^G = \theta$. These minima are attainable since they belong to the polytope defined by (12) relative to the correct decision. The cost functions then present minimum values equal to: $J_d^{min} = J_c^{min} = 0$.

We have shown in the last subsection that the DD cost function presents different expressions (equation (16)) according to the considered polytope among three polytopes corresponding to incorrect decisions. Consequently, in each polytope, the DD cost function, on the contrary of the CMA cost function, offers a different minimum:

$$\hat{\theta}_d^L = \begin{cases} -1 & \text{if} \quad \hat{\theta} < -1 - |\theta|, \\ 0 & \text{if} \quad |\hat{\theta}| < -1 + |\theta|, \\ +1 & \text{if} \quad 1 + |\theta| < \hat{\theta}; \end{cases} \tag{17}$$

$$\hat{\theta}_c^L = 0.$$

Note that these values are independent of the channel parameter θ.

These minima correspond to possible local minima of the both criteria. Indeed, the only attainable local minimum is here $\hat{\theta}_d^L = \hat{\theta}_c^L = 0$ occurred in the case of maximum phase channels. The other values (obtained for the DD criterion) are not included in the respective polytopes.

Note that the local minima here enumerated correspond only to the regions of incorrect decisions such as $E\{\hat{a}_n a_n\} = 0$. Let's recall that it exists also a region of uncertain decisions such as $E\{\hat{a}_n a_n\} \neq 0$ and $E\{\hat{a}_n a_n\} \neq 1$ which is not studied and which can possibly present a second local minimum.

IV. SOFT DECISION SOLUTION

The investigation of the minima done in the last section has shown that the error surface consists in polytopes separated by hard frontiers sometimes difficult to cross by the blind equalization algorithm. This is due to the nonlinearity used in the decision device and so to the fact that data belong to a finite alphabet.

In this section, we first analyze the effect of the introduction of the soft decision on the error surface associated with the blind decision-directed and constant modulus equalization criteria. We also show how the soft decision, once used during a phase of initialization called the pseudo-training phase, allows the corresponding algorithms to escape from the local minimum and to converge to the desired minimum ensuring a correct equalization.

A. Modification of the Mean Error Surface

The existence of local minima is linked to the discrete character of the transmitted data a_n and of the consequently reproduced data $\hat{a}_n$. This restitution can be made only by the use of a hard decision that introduces a discontinuity in the error surface. This involves, as we have seen in the last Section, the creation of small portions of surface separated by hard frontiers sometimes difficult to cross by the blind equalization algorithm with possible attainable local minima.

The idea that we expose to prevent the blind algorithm from convergence to the local minima, resides in the preliminary use of a soft nonlinear function in the decision device. We propose to use the tangent hyperbolic function characterized by the saturation factor s, defined as:

$$Q(x) = \tanh\left(\frac{x}{s}\right). \tag{18}$$

Fig. 2 exhibits the error surface associated with the DD criterion, for different values of s. This figure shows that the curves, corresponding to the utilization of the soft decision, present transitions (frontiers between the different polytopes) smoother than those corresponding to the hard decision (implemented by the function $\text{sgn}(.)$ since we use $\hat{a}_n = \pm 1$). This phenomenon is emphasized as the factor of saturation increases that is to say as we distance the hard decision. The curve corresponding to $s = 0.1$ presents almost the same cavity and the same transition levels as those corresponding to the case of the hard decision, while the curve corresponding to $s = 1.0$ presents a flat cavity with no transition between the polytopes of the local minimum and the global minimum.
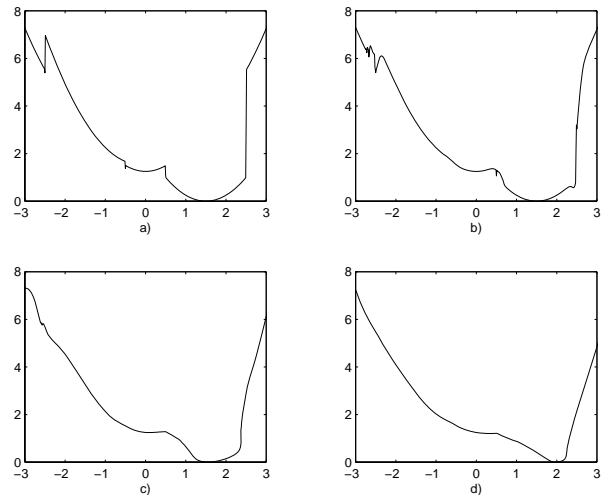

Fig. 2 Mean error surface of the DD cost function $J(\hat{\theta})$:

a) hard decision case, b), c) and d) soft decision cases respectively for s = 0.1, 0.6 and 1.0 (maximum phase channel: $\hat{\theta} = 1.5$)

However, the portion of the error surface corresponding to the global minimum presents this time a deeper cavity. The curve relative to $s = 0.1$ presents a shape similar to that relative to hard decision: a flat cavity between the two-polytope limits. That relative to $s = 1.0$, presents a very pronounced cavity around the global minimum.

Fig. 2 also allows us to observe that the global minimum is no longer the same according to whether one introduces or not the soft decision. Indeed, in the case of the soft decision, the global minimum $\hat{\theta}_{opt}$ shifts according to s. It is 1.5 in the case

of the hard decision, while for the two values of the factor of saturation, mentioned higher, it is respectively 1.51 and 1.78. The displacement of the global minimum is small for the small values of s.

We then observe, in this example, that the error surface is modified. The transitions between polytopes are smoothed. The local minimum is attenuated and the global minimum is displaced. This allows us to anticipate the interest of the use of the soft decision in the adaptive case: the blind algorithm, once trapped in a local minimum, may escape from it more easily than in the case of the hard decision.

B. Modification of the convergence properties

The use of the soft decision in the decision device results in replacing the vector $\hat{A}_n$, present in the adaptation algorithms DD and CMA (equations (4) and (6)), by:

$$\tilde{A}_n = \left(\tilde{a}_{n-1}, \tilde{a}_{n-2}, \cdots, \tilde{a}_{n-M+1} \right); \qquad (19)$$

where $\tilde{a}_n = Q(c_n)$.

The analysis above shows that more the saturation factor increases more the global minimum moves over the desired global minimum. Then, we here suggest using a soft decision only at the beginning of the DFE adaptation, i.e. during a so-called pseudo-training phase. The use of the soft decision during the pseudo-training phase allows the algorithm to converge to a new minimum presenting a value near the global minimum. Once the algorithm has converged to this minimum, we can switch to the hard decision, that is, to a properly tracking phase, to reach the desired global minimum and to guarantee the correct equalization of the channel.

This approach is illustrated by the following example presented in Fig. 3.a. The DD algorithm is performed at the beginning with the soft decision during 500 samples, at the end of which the algorithm approaches the value 0.76: $s = 0.6$, $\hat{\theta}(0) = 0$ and $\mu = 0.001$ (see the Fig. 2.c for the corresponding error surface). After the switching to the hard decision (end of the pseudo-training phase), the algorithm hangs on to the global minimum $\hat{\theta} = \theta = 1.5$. This curve is compared to that of the DD algorithm, which is still trapped in the local minimum for the same value of the adaptation step size. The corresponding mean square error (MSE $= E\left\{ \left(a_n - c_n \right)^2 \right\}$) evolution is presented in Fig. 3.b. Fig. 3.c and Fig. 3.d present the same illustrations relatives to the CMA equalizer.

C. New algorithms based on the use of soft decision

According to the foregoing, to smooth and then avoid the local minima, we must choose a high value for the saturation factor during the pseudo-training-phase. However, as s increases the equalizer parameter $\hat{\theta}(n)$ goes away from the global minimum at the convergence. This may distort the equalization even, if we switch later to the hard decision.

So we can conclude that we have to choose a high value of the saturation factor at the initialization of the algorithm to avoid the local minima and a low value at the convergence such that the minima corresponding to the two kinds of

decisions (hard and soft) are close enough to ensure convergence to the desired global minimum. To make variable the duration of the pseudo-training-phase according to the algorithm behavior towards a given channel, we propose to use the gradient based algorithm, minimizing the mean square error $E\left\{ \left(\hat{a}_n - c_n \right)^2 \right\}$, to calculate the evolution of s. The relative adaptation equation is written:

$$s(n+1) = s(n) + \mu \left(\hat{a}_n - c_n \right) \Gamma \left(s(n) \right) \qquad (20)$$

where: $\Gamma \left(s(n) \right) = \nabla_s \left. \left(\hat{a}_n - c_n \right) \right|_{s=s(n)}$.

The expression of $\Gamma \left(s(n) \right)$ is deduced from the (2), (18) and (19):

$$\Gamma \left(s(n) \right) =$$
$$\sum_{i=1}^{L} \left(\frac{\hat{\theta}_i(n)}{s(n)} \right) \left[1 - \tanh \left(\frac{c_{n-i}}{s(n)} \right)^2 \right] \left(\frac{c_{n-i}}{s(n)} - \frac{\partial c_{n-i}}{\partial s(n)} \right). \qquad (21)$$

Account held of the recursive form of $\Gamma \left(s(n) \right)$, the exact calculus of this term will necessitate then an infinite memory. Simulations that we have undertaken have shown that the next approximation:

$$\Gamma(s(n)) = \sum_{i=1}^{L} \left(\frac{\hat{\theta}_i(n)}{s(n)} \right) \left[1 - \tanh \left(\frac{c_{n-i}}{s(n)} \right)^2 \right] \left(\frac{c_{n-i}}{s(n)} \right) \qquad (22)$$

is acceptable. This is equivalent to the approximation version of the Least Mean Square algorithm (LMS): the Extended LMS [14], when applied to a recursive filter. This approximation supposes that $s(n)$ varies slowly.

To summarize, let's say that the modified blind equalization algorithms: modified DD a modified CMA, are those achieving an automatic switch between the initialization phase (during which we use a soft decision) and the properly tracking phase (during which we use a hard decision). The implementation of these modified algorithms will be done by taking into account (4) or (6) (adaptation of the equalizer impulse response), (18) and (19) (use of a soft decision). So, the modified DD algorithm is written:

$$\hat{\Theta}_d(n+1) = \hat{\Theta}_d(n) - \mu(\hat{a}_n - c_n)\tilde{A}_n; \qquad (23)$$

where $\tilde{A}_n = \left(\tilde{a}_{n-1}, \tilde{a}_{n-2}, \cdots, \tilde{a}_{n-M+1} \right)$, $\tilde{a}_n = \tanh \left(c_n / s(n) \right)$.

The modified CMA is written as:

$$\hat{\Theta}_c(n+1) = \hat{\Theta}_c(n) + \mu c_n (c_n^2 - \gamma)\tilde{A}_n. \qquad (24)$$

The adaptation of the saturation factor $s(n)$ is done according to (20).

V. COMPUTER SIMULATIONS

In the following, to prove the efficiency of the modified algorithms using the soft decision during the pseudo-training-phase, we consider first noisy channels: Proakis A and Proakis B channels [15]. The SNR is fixed at 10 dB. We consider also a complete DFE with a forward filter of 21 coefficients and a feedback filter of 5 coefficients.

To show the capacity of the soft decision to help the DFE to avoid the ill-convergence and consequently to reach the desirable functioning, we have initialized the coefficients of

the forward filter and the feedback filter of the DFE by two times in addition to the initial initialization.

The results of the simulation under the conditions mentioned above are presented in the Fig. 4, where only the CMA is used. In this figure, we illustrate the evolution versus time of the output signal of the DFE (input of the decision device) for the two channels. In this figure, we present a comparison between the behavior of the DFE without (curves 1.a and 2.a respectively) and with the soft decision (curves 1.b and 2.b respectively). The evolution versus time of the saturation factor s, is also presented for the two channels in the curves 1.c and 2.c respectively. For all this curves the DFE is initialized at the times: $n = 0, 1000, 2000$.
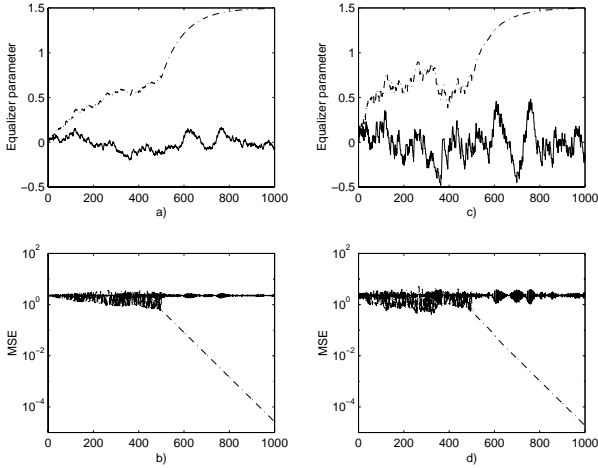

Fig. 3 Evolution, without (in continuous lines) and with the soft decision (in discontinuous lines), of $\hat{\theta}(n)$ and MSE versus time: curves a) and b) DD algorithm, and curves c) and d) CMA ($\hat{\theta} = 1.5$)

It is obvious from the Fig. 4 (case of Proakis A channel) that the DFE succeed to reach the desirable functioning more rapidly when the soft decision is used. Moreover, the DFE is unable in certain situations (case of Proakis B channel) to avoid the ill-convergence if the soft decision is not used. This curves shows the interest to use the soft decision in the initialization phase or when the equalizer deviate from the desirable functioning (due to variation of the channel behavior, finite-precision design and implementation of the DFE, ...).

VI. CONCLUSIONS

In this paper, we have presented the behavior of the decision feedback equalizers adapted by the decision-directed algorithm or the constant modulus algorithm. The blind algorithm error surface analysis has inspired us to replace the hard limiter in the decision device by soft decision during the initialization of the blind algorithms. This was shown to reduce the influence of false decisions and to smooth the local minima so that the convergence to the desired minimum may occur for any step-size and any initialization. We have then proposed to modify the classic DD algorithm and the CMA by using a soft decision implemented by a tangent hyperbolic whose slope is adapted according to the error decision.

Though the theoretical analysis presented in this paper concerns the noiseless case, we have shown by simulation the robustness of the soft decision DFE with respect to noisy transmission channel.

The use of the soft decision could be extended to any blind algorithm since the problem of the discrete character of the data and so the hard transitions between the portions forming the error surface cost function, appears by the same manner.
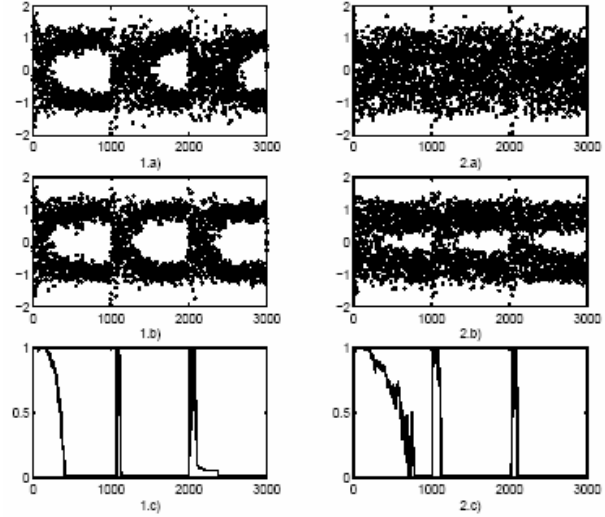

Fig. 4 DFE behavior in presence of channels Proakis A (1) and B (2), SNR = 10 dB. 1.a and 2.a: output without soft decision, 1.b and 2.b: output with soft decision, 1.c and 2.c: evolution of s(n)

APPENDIX

Appendix A

The goal is the calculus of the term $E\{a_n \hat{a}_{n-1}\}$. Let's determine the expression of $a_n \hat{a}_{n-1}$:

$$
\begin{aligned}
a_n \hat{a}_{n-1} &= a_n \mathrm{sgn}(c_{n-1}), \\
&= a_n \mathrm{sgn}(a_{n-1} + \theta a_{n-2} - \hat{\theta}\hat{a}_{n-2}).
\end{aligned}
\tag{25}
$$

Since we consider the incorrect decision case, the term $\mathrm{sgn}(a_{n-1} + \theta a_{n-2} - \hat{\theta}\hat{a}_{n-2})$ is independent of the value of a_{n-1}, then:

$$
\begin{aligned}
a_n \hat{a}_{n-1} &= a_n \mathrm{sgn}(\theta a_{n-2} - \hat{\theta}\hat{a}_{n-2}), \\
&= \mathrm{sgn}(\theta a_n a_{n-2} - \hat{\theta} a_n \hat{a}_{n-2}).
\end{aligned}
\tag{26}
$$

We have also used above the fact that a_n takes the values ± 1 and so $a_n = \mathrm{sgn}(a_n)$.

The term $a_n a_{n-2}$ takes the values ± 1 with equal probability since the data $\{a_n\}$ are iid and $Pr\{a_n = 1\} = Pr\{a_n = -1\} = 1/2$. In addition, a_{n-2} takes the values $\pm \hat{a}_{n-2}$ with equal probability (incorrect decisions:

$E\{a_{n-2}\hat{a}_{n-2}\} = 0$), and so $a_n a_{n-2}$ takes the values $\pm a_n \hat{a}_{n-2}$ with equal probability (i.e. $Pr\{a_n a_{n-2} = a_n \hat{a}_{n-2}\} = Pr\{a_n a_{n-2} = -a_n \hat{a}_{n-2}\} = 1/2$).

Consequently, the term $a_n \hat{a}_{n-1}$ takes the values $\pm(\theta + \hat{\theta})$ and $\pm(\theta - \hat{\theta})$ with equal probability (i.e. with a probability equal to 1/4). We deduce then that $E\{a_n \hat{a}_{n-1}\} = 0$.

Appendix B

The goal is the calculus of terms $E\{\hat{a}_n a_{n-1}\}$ and $E\{\hat{a}_n \hat{a}_{n-1}\}$. By using the same assumptions that in appendix A, we write:

$$
\begin{aligned}
\hat{a}_n a_{n-1} &= \mathrm{sgn}(c_n) a_{n-1}, \\
&= \mathrm{sgn}\left(a_n + \theta a_{n-1} - \hat{\theta}\hat{a}_{n-1}\right) a_{n-1}, \\
&= \mathrm{sgn}\left(\theta a_{n-1} - \hat{\theta}\hat{a}_{n-1}\right) a_{n-1}, \\
&= \mathrm{sgn}\left(\theta - \hat{\theta} a_{n-1}\hat{a}_{n-1}\right).
\end{aligned}
\tag{27}
$$

The term $a_{n-1}\hat{a}_{n-1}$ takes the values ± 1 with equal probability since we consider here the incorrect decision case ($E\{a_{n-1}\hat{a}_{n-1}\} = 0$) and so $Pr\{\theta - \hat{\theta}a_{n-1}\hat{a}_{n-1} = \theta - \hat{\theta}\} = Pr\{\theta - \hat{\theta}a_{n-1}\hat{a}_{n-1} = \theta + \hat{\theta}\} = 1/2$.

Now, let's determine the sign of $\theta - \hat{\theta}$ and $\theta + \hat{\theta}$ according to the three polytopes corresponding to the incorrect decision case (equation (17)):

polytope 1 :

$\hat{\theta} < -1 - |\theta| \quad \rightarrow \quad \mathrm{sgn}(\theta - \hat{\theta}) = 1, \mathrm{sgn}(\theta + \hat{\theta}) = -1$

polytope 2 : $\qquad (|\theta| > 1)$

$|\hat{\theta}| < -1 + |\theta| \quad \rightarrow \quad \mathrm{sgn}(\theta - \hat{\theta}) = \mathrm{sgn}(\theta + \hat{\theta}) = \mathrm{sgn}(\theta)$

polytope 3 :

$\hat{\theta} > 1 + |\theta| \quad \rightarrow \quad \mathrm{sgn}(\theta - \hat{\theta}) = -1, \mathrm{sgn}(\theta + \hat{\theta}) = 1.$

$$\tag{28}$$

Consequently, we can deduce that:

$$
E\{\hat{a}_n a_{n-1}\} = \begin{cases} 0, & \text{if} \quad \hat{\theta} < -1 - |\theta|; \\ \mathrm{sgn}(\theta), & \text{if} \quad |\hat{\theta}| < -1 + |\theta| \text{ and } |\theta| > 1; \\ 0, & \text{if} \quad \hat{\theta} > 1 + |\theta|. \end{cases}
\tag{29}
$$

By the same manner, we can calculate $E\{\hat{a}_n \hat{a}_{n-1}\}$:

$$
\begin{aligned}
\hat{a}_n \hat{a}_{n-1} &= \mathrm{sgn}(c_n) \hat{a}_{n-1}, \\
&= \mathrm{sgn}\left(a_n + \theta a_{n-1} - \hat{\theta}\hat{a}_{n-1}\right) \hat{a}_{n-1}, \\
&= \mathrm{sgn}\left(\theta a_{n-1} - \hat{\theta}\hat{a}_{n-1}\right) \hat{a}_{n-1}, \\
&= \mathrm{sgn}\left(\theta a_{n-1} \hat{a}_{n-1} - \hat{\theta}\right).
\end{aligned}
\tag{30}
$$

The term $\theta a_{n-1}\hat{a}_{n-1} - \hat{\theta}$ takes the values $\theta - \hat{\theta}$ ($a_{n-1}\hat{a}_{n-1} = 1$) or $-\theta - \hat{\theta}$ ($a_{n-1}\hat{a}_{n-1} = -1$) with equal probability. The sign of

these two values are calculated before since $\mathrm{sgn}(-\theta - \hat{\theta}) = -\mathrm{sgn}(\theta + \hat{\theta})$. Finally, we can deduce that:

$$
E\{\hat{a}_n \hat{a}_{n-1}\} = \begin{cases} 1, & \text{if} \quad \hat{\theta} < -1 - |\theta|; \\ 0, & \text{if} \quad |\hat{\theta}| < -1 + |\theta| \text{ and } |\theta| > 1; \\ -1, & \text{if} \quad \hat{\theta} > 1 + |\theta|. \end{cases}
\tag{31}
$$

REFERENCES

[1] Z. Ding, Y. Li, and K. D. Ding, "*Blind Equalization and Identification*.," Marcel Dekker, 2001.

[2] K. Pahlavan and A. Levesque, "*Wireless Information Networks*," John Wiley & Sons, 1995.

[3] A. Benveniste, M. Goursat, and G. Ruget, "Robust identification of a nonminimum phase system : Blind adjustment of a linear equalizer in data communications," *IEEE Transactions on Automatic Control*, vol. 25, no. 3, pp. 385–398, June 1980.

[4] O. Macchi and E. Eweda, "Convergence analysis of self-adaptive equalizers," *IEEE Transactions on Information Theory*, vol. 30, no. 3, pp. 161–176, Mar. 1984.

[5] O. Macchi and A. Hachicha, "Self-adaptive equalization based on a prediction principle," in *IEEE Global Telecommunication Conference*, Dec. 1986, pp. 1641–1645.

[6] E. Bai and Z. Ding, "Blind decision feedback equalization of timevarying channels with DPSK inputs," *IEEE Transactions on Signal Processing*, vol. 49, no. 7, pp. 1533–1542, July 2001.

[7] A. Rontogiannis and K. Berberidis, "Efficient decision feedback equalization for sparse wireless channels," *Efficient Decision Feedback Equalization for Sparse Wireless Channels*, vol. 2, no. 3, pp. 570–581, May 2003.

[8] Y. Li and Z. Ding, "Global convergence of fractionally spaced godard adaptive equalizers," *28th Asilomar COnference on Signals, Systems and Computers*, pp. 617–621, Nov. 1994.

[9] I. Fijalkow, C. E. Manlove, and J. C. R. Johnson, "Adaptive fractionally spaced blind cma equalization: Excess mse," *IEEE Transactions on Signal Processing*, vol. 46, no. 1, pp. 227–231, Jan. 1998.

[10] Z. Ding, R. Kennedy, B. Anderson, and C. Johnson, "Local convergence of the sato blind equalizer and generalizations under practical constraints," *IEEE Transactions on Information Theory*, vol. 39, no. 1, pp. 129–144, Jan. 1993.

[11] R. Kennedy, B. Anderson, and R. Bitmead, "Blind adaptation of decision feedback equalizers : Gross convergence properties," *International Journal of Adaptive Control and Signal Processing*, vol. 7, pp. 497–523, 1993.

[12] D. Godard, "Self-recovering equalization and carrier tracking in twodimensional data communication systems," *IEEE Transactions on Communications*, vol. 28, no. 11, pp. 1867–1875, Nov. 1980.

[13] R. Kennedy and B. Anderson, "Recovery times of decision feedback equalizers on noiseless channels," *IEEE Transactions on Communications*, vol. 35, no. 10, pp. 1012–1021, Oct. 1987.

[14] P. Feintuch, "An adaptive recursive LMS filter," *Proc. IEEE*, pp. 1622–1624, 1976.

[15] J. Proakis, "*Digital Communications*," McGraw-Hill, 2000.