Decision Feedback Blind Equalizer: Information-Maximization Approach

Vladimir R. Krstic, Member, IEEE, and Zoran Petrovic, Member, IEEE

Abstract — This paper presents an innovative solution of a blind decision feedback equalizer (DFE) based on information-maximization approach, i.e., joint entropy maximization criterion (JEM), where a “soft” non-linearity incorporated into the decision device transforms the probability density function of the input sequence. This approach applied in two-dimensional communication system induces the problem of a non-linear complex mapping of estimated symbols. Using Taylor series extension of the hyperbolic tangent function and some simplifications of the JEM algorithms to complex plane and to preserve desirable activation properties of these algorithms.

Keywords — Blind decision feedback equalization

I. INTRODUCTION

THE specific problem of the blind activation of the decision feedback equalizer (DFE) is the phenomenon of error propagation, which takes place at the moment of transition from the blind acquisition to conventional decision directed (DD) mode. This problem is critically emphasized with the transmission channels introducing the severe intersymbol interference (ISI) (e.g., multipath radio or underwater acoustic channels). In this propagation environment the well known and widely used blind algorithms designed for linear equalizer, for instance Constant Modulus Algorithm (CMA) [1], are not able to open the eye diagram enough and the error propagation certainly produces an activation fail. On the other hand, the “self-adaptive” DFE [2] is the method originally designed to overcome this problem. This type of blind activation is based on the ability of DFE to self-adapt both its structure and its adaptation criterion, according to some performance measure. The basic idea of this solution is to divide the difficult task of the blind activation into easier subtasks: in the blind acquisition mode the equalizer is transformed into the cascade of linear and decoupled devices, and in the tracking mode it switches into the conventional DD DFE, which minimizes the mean square error (MSE) using LMS algorithm, Fig. 1. The most important characteristic of this transformation is the position of the key components: (1) the decorrelator ($R$) before linear equalizer ($T$) in the blind acquisition, and (2) the $R$ after $T$ in the phase of tracking. This position swapping is based on an assumption that the coefficients setup of the pure recursive whitening filter $R$ at the moment of structure switching is what the feedback (FBF) part of DFE needs in the tracking mode. Unfortunately, the positions of $R$ and $T$ are irrelevant only in the steady state but it becomes critical in the blind mode. As a consequence, the inadequate initial coefficients setup of the FBF realized by the “hard” decision device (slicer) (see Fig.1) will certainly activate error propagation in the case of severe channels. The Soft-DFE solution presented in [3], [4] is an example how this difficulty can be mitigated: (1) using a more efficient decorrelation algorithm than the Extended LMS [2] to improve blind acquisition, and also (2) introducing a new “soft” transition mode. The principal feature of this solution are new algorithms for a decorrelation and soft transition mode, both based on the joint entropy maximization (JEM) criterion [5], [6].

The theoretical framework for introducing JEM as a criterion function for equalizer optimization is the information-maximization concept [5], where the minimization of ISI is perceived as a problem of redundancy reduction of the output sequence. This approach has been used in [6] to develop the class of JEM stochastic gradient algorithms for DFE activation supposing real-valued signals and one-dimensional system. Unfortunately, the direct extension of these algorithms to complex counterparts is not possible due to problem of non-linear mapping of complex-valued symbols. This paper addresses new aspects of using the JEM criterion in the context of two-dimensional data transmission system.

In Section II we discussed the extension problem of real-valued version of the JEM algorithms to complex counterparts in the context of the Soft-DFE solution. The simulation results in Section III present the performance of the JEM algorithm in the transition mode.
II. COMPLEX JEM ALGORITHMS FOR SOFT-DFE

A. Problem definition

Consider the time-discrete system with the channel and the equalizer in Fig. 2a, [6]. The BPSK symbols \( a(k) \) feed the system and its output, \( x(k) \), at the moment \( t = kT \) (\( T \) is symbol interval), is the mixture of current and previous symbols that have been distorted by time invariant linear channel. Time delay introduced by system is irrelevant in this consideration. The reduced structure of DFE, which has only the FBF, simplifies an analysis but it is not an obstacle since it is possible to separate the adaptation of the FBF from the feedforward (FFF) part. A classical slicer is replaced with a soft function \( g(\cdot) \) that must be non-linear, monotone increasing and memory-less. This activation function should transform, in an iterative manner, the input sequence with an unknown probability density function to a maximum-entropy sequence with the uniform density. For instance, the hyperbolic tangent function, which is often used as an activation (sigmoidal) function in neural networks [7], has desirable characteristics for non-linear mapping in DFE: bounded at \( \pm \alpha \), and approximately linear, with greatest slope at the axis crossing.

The gradient descent algorithm derived in [6], which maximizes the joint entropy of the equalizer output, is

\[
 b_n(k+1) = b_n(k) + \mu E \left[ \frac{\partial \ln | J |}{\partial b_n(k)} \right] 
\]

where \( \mu \) is a step size and \( |J| = \frac{\partial r(k)}{\partial y(k)} \) is absolute value of the Jacobian transformation of the output sequence \( r_1(k),...,r_{N+1}(k) \), \( N \) is the length of FBF filter. The above algorithm is derived under the assumption that all previous decisions are correct, i.e., \( r(k-n) = a(k-n) \), \( n = 1,...,N \), (an assumption commonly used in DFE analysis). In the case when symbol estimates \( y(k) \) are mapped by the hyperbolic tangent function, \( g(y(k)) = \alpha \tanh[\beta y(k)] \), the stochastic approximation of gradient in (1) is

\[
 \frac{\partial \ln | J |}{\partial b_n(k)} = \frac{\partial \ln | J |}{\partial b_n(k)} \bigg|_{y(k)} = -2\alpha \tanh[\beta y(k)] r(k-n) 
\]

\[
 = -2\frac{\beta}{\alpha} r(k) r(k-n). 
\]

Combining (2) into (1) and omitting mathematical expectation \( E \) the basic JEM algorithm is derived

\[
 b_n(k+1) = b_n(k) - \mu r(k-n). 
\]

Thus, the instantaneous gradient in (3) is a function of the nonlinearity \( \alpha \tanh[\beta y(k)] \) applied in the decision device.

B. Complex JEM algorithms

Derivation of the complex JEM algorithms is based on approximation of \( \tanh \) function and modification of JEM-DFE scheme. These simplifications should provide a best tradeoff between the implementation complexity of the decision device and a necessity to preserve characteristics found in the real-valued \( \tanh \) function. In that sense, following approximations are possible:

1. Using Taylor series expansion of \( \alpha \tanh[\beta y(k)] \), where only the first two terms are considered, the algorithm (3) is transformed in

\[
 b_n(k+1) = b_n(k) - \mu_B y(k) \left[ 1 - \beta_B y^2(k) \right] r(k-n) 
\]

where \( \mu_B = \mu_\alpha \beta \) is a step size, and \( \beta_B = \frac{\beta^2}{3} \) is parameter (fits the slope of nonlinearity at the axis.

Note, the parameters \( \alpha \) and \( \beta \) can be used as a tool to vary the algorithm characteristics.

Let us now consider the JEM algorithm (3) and the basic JEM-DFE scheme in a context of a commonly used two-dimensional signals (M-PSK and M-QAM). The specific task that is emerged here is the extension of real-valued non-linearity to the complex plane. The extension of the hyperbolic tangent function to complex plane with the principle of analytical continuation does not imply continuity of behavior between a real-valued function and its extension to the complex plane [7]. In other words, when the real component of \( \tanh \) approaches zero \( \tanh \) is not bounded and tends rapidly to infinity as the imaginary part approaches \( \pm \pi/2 \):

\[
 \tanh(y) = \tanh(y' + iy'') \\
 \text{when } y' \to 0 \\
 \tanh(y) = \tanh(iy') = i \tan(y') 
\]

This behavior is critical and suggests carefullness with the extension of real-valued JEM algorithm to the corresponding complex solution. Even though, it does not mean that this function is entirely unusable as an activation function in the two-dimensional DFE.

Fig 2. System with JEM-DFE: (a) basic JEM-DFE scheme, and (b) modified JEM-DFE scheme
The second modification is a step size, where we replaced the soft decisions \( r(k-n) \) with the outputs of the slicer, see Fig. 2(b). In this modified JEM-DFE scheme the estimates of the transmitted symbols \( \hat{a}(k-n) \) feed the FBF filter. The algorithm (5) becomes

\[
b_n(k+1) = b_n(k) - \mu_D y(k) \left[ 1 - \beta_2 y^2(k) \right] \hat{a}(k-n) \tag{6}
\]

The above algorithm is a practical approximation of (3).

Considering a basic JEM-DFE scheme it is realistic to expect that the complex-valued inputs to the sigmoid reaches conditions described by (4) in the case of very severe channels. On the other hand, with the modified JEM-DFE scheme the probability of this event is neglectable since the real and imaginary components of the complex-valued symbols at the output of the slicer are different from zero. As a consequence, the real part of weighted sum of previous decisions, i.e., output of FBF filter moves away from zero. This analysis indicates that we can extend the real-valued algorithm (6) to the complex one using the analogy with the complex LMS algorithm [8]. Thus, the complex JEM-D algorithm [3] is

\[
b_n(k+1) = b_n(k) - \mu_D, y(k) \left[ 1 - \beta_2 y^2(k) \right] \hat{a}^*(k-n) \tag{7}
\]

where \( \mu_D \) is a step size, and \( \beta_2 \) is the parameter of JEM-DFE. The variables \( b_n(k), y(k), \) and \( \hat{a}(k) \) are complex, and the symbol * denotes complex-conjugation.

It is clear that the above algorithm must be able to adapt simultaneously both the real and the imaginary part of \( b_n(k) \) by maximizing the joint entropy of the real and the imaginary parts of \( y(k) \left[ 1 - \beta_2 y^2(k) \right] \hat{a}(k-n) \). Thus, the overall effect of transformations from (3) to (7) should be similar to what might be expected from a true “complex sigmoid”.

Finally, besides the algorithm (7), the Soft-DFE has another JEM type algorithm. Consider again “self-adaptive” structure in Fig. 1. As we indicated in Section I, the positions of \( R \) and \( T \) are irrelevant in the steady state, i.e., \( R \) can be placed after or before \( T \). Thus, depending on these positions \( R \) obtains the decision device (recursive DFE) or loses it (linear pure recursive filter). This transformation of the cascade is followed by corresponding modification of algorithm (7). In the case of decorrelator \( R \) the algorithm becomes JEM-W [3]

\[
b_n(k+1) = b_n(k) - \mu_W y(k) \left[ 1 - \beta_1 y^2(k) \right] y(k-n) \tag{8}
\]

where \( \mu_W \) is a step size, and \( \beta_1 \) is the parameter.

Even though this decorrelation algorithm is a result of a formal extension of JEM-D, it really performs much better (see [3], [4]) in comparison with the Extended LMS algorithm [2].

C. Structure of Soft-DFE

The Soft-DFE is a T/2 fractionally-spaced equalizer (FSE) that performs three operation modes: blind acquisition, soft transition and tracking. During the acquisition mode the \( GC \cdot R \cdot T \) cascade is linear equalizer where \( GC \) is gain control, \( R \) is the JEM-W decorrelator and \( T \) is the FSE using CMA algorithm. The decorrelator consists of two decoupled all-pole whitening filters. The coefficient vector of one of whiteners is translated into the FBF part of DFE after structure switching. In the soft transition mode, the joint equalization and carrier phase tracking take place so that the linear part is now the DD LMS equalizer and the nonlinear part is the decision feedback that maximizes joint entropy, i.e., the modified JEM-DFE.. Note that \( R \) is suppressed after this switching. The tracking mode is the same as in a classical MMSE DFE case. The switching control of different operation modes is provided by a performance monitoring device. This MSE estimator monitors two thresholds: \( M_{TL-1} \) from acquisition to soft transition, and \( M_{TL-2} \) from soft transition to tracking mode. The detailed description of the Soft-DFE is not subject of this paper. (Please, refer to [3] and [4] for the in-depth look of this DFE solution.)

III. SIMULATION RESULTS

In this section we present the results of testing Soft DFE activation, or more precise the performance of the transition mode, which is dominantly dependent on JEM-D algorithm (7). The applied software simulator is the upgraded variant of the V.32-modem simulator. All these results are carried out using Monte Carlo tests with 1000 independent runs and channels with signal-to-noise ratio SNR=25 dB. The transfer function of applied transmitter and receiver filters follows a raised cosine with roll-off factor 0.12. The channels represent a three-ray multipath environment whose impulse response is given by

\[
b(t) = e(t)W(t) + d_1e(t-\tau_1)W(t-\tau_1) + d_2e(t-\tau_2)W(t-\tau_2) \tag{9}
\]

where \( e(t) \) is the basic pulse, \( W(t) \) is a rectangular window spanning [-16T,16T], \( d_i \) is attenuation factor of \( i \)th channel, and \( \tau_i \) is propagation delay of \( i \)th path [6]. The multipath parameters \( d_1 \) and \( \tau_1 \) take the following values for channels B, C, D, E, F: \( d_1 \in \{0.8,0.9,0.8,0.9,0.7\} \), \( d_2 \in \{0.35,0.35,0.4,0.4,0.4\} \), \( \tau_1 \in \{2q,3q,2q,3q,2q\} \), respectively, where \( q = 16/4 \) \( \tau_2 = 2T \). The amplitude response characteristics of the combination (transmitter filter + channel) for these channels are given in Fig. 3. The impulse response spans of two parts, FFF and FBF, are 22 T and 6T, respectively. The initial weights of two centered reference taps (two spike strategy) of the FSE are 2.4. The thresholds levels of both, the structure and the algorithms switching, are experimentally defined for 16-QAM and 32-QAM: \( M_{TL-1} = 1.5 dB \) and \( M_{TL-2} = -8dB \).
The testing of overall effects of the JEM-D algorithm with respect to error propagation and convergence of MSE in transition mode is based on testing the impact of the smoothing parameter $\beta_2$. It is used to vary the slope of the activation function. In this way we have also proved the optimization possibility of this parameter with respect to the statistic of signal constellation (defined by the constant $R_2$). Fig. 4 represents duration of the transition mode, expressed in T intervals, versus $\beta_2$, for different channels, the signals 16-QAM and 32-QAM, and the fixed $\beta_1 = 1$. The transition mode is bounded by the thresholds $M_{TLM-1}$ and $M_{TLM-2}$. Fig. 5 represents the corresponding symbol error rate, obtained by normalization with respect to the duration of the transition mode which was reached for $\beta_2 = 2$.

The presented results shows that the optimal value of $\beta_2$ may be find for the known statistical characteristics of the transmitted signal. In the other wards, the performance characteristics of the JEM-D dominantly depend on the signal statistic and very little on the channel. The Fig. 6 indicates that the optimal value of $\beta_2$ with respect to error propagation effects, $\beta_2 \approx 6$, shows also optimality for MSE convergence.

IV. CONCLUSION

The presented approach of the blind decision feedback equalization based on innovated joint entropy algorithms gives an opportunity to improve robustness with respect to error propagation effects. Using a smoothing parameter it is possible to maximize joint entropy of output sequence, i.e. to minimize ISI. The optimal value of this parameter is dominantly dependent on the statistic of the applied symbol constellation, what strongly confirm a practical worth of the presented Soft-DFE.

REFERENCES