Complex-Valued Maximum Joint Entropy Algorithm for Blind Decision Feedback Equalizer

Vladimir R. Krstić[1], Zoran Petrović[2]

Abstract - **This paper proposes the complex domain solution of the blind decision feedback equalizer (DFE) based on joint entropy maximization (JEM) criterion. For the feedback filter of DFE, a new complex activation function is defined and then the stochastic gradient algorithm of the JEM type is derived. The fact that the JEM cost is a real function of complex quantities suggests the method associated to a least man-square criterion.**

Keywords – **Blind equalization, joint entropy maximization**

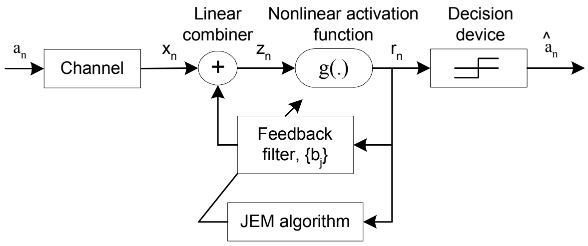

Fig. 1. Model of transmission system with soft feedback filter

I. Introduction

The real to complex domain extension is a problem that, in general, characterizes the usage of neural networks [1]. The most of real-valued nonlinearities, which have been commonly used as information-theoretic objective functions (activation functions), become unusable in the complex domain due to difficulties following their extension into the complex domain. In fact, the conflict between the boundedness and the analyticity of complex nonlinearities causes a lack of appropriate complex activation functions.

In this paper we consider and propose the complex domain solution of the blind *decision feedback equalizer* (DFE) that was derived in [2] for real-valued input data by using the information-maximization theoretic approach [3]. The basic set of suitable properties of a complex activation function, which is defined in the context of the multilayer feedforward neural networks (multilayer perceptron) [4]-[6], is modified in order to support the DFE adaptation based on the joint entropy maximization (JEM) criterion. Hence, for a new complex activation function, the complex-valued stochastic gradient algorithm of the JEM type (CJEM) is derived following the method applied to the multilayer perceptron in [4] and [5].

II. Problem Definition

In [2], the feedback part of the DFE is considered as a sigmoid unit in the system with a real channel and binary symbol source $\{a_n\}$ that is independent, identically distributed (i.i.d.) and zero mean, Fig 1. In the this unit, which is denoted as a soft feedback filter (SFBF), the output of the linear combiner, z_n, is the sum of channel output x_n and weighted sum of previous outputs r_{n-j}, $j = 1,...,N$, feeding the delay line of the feedback filter defined by coefficients $\{b_j\}$. The real-valued activation function $g(\cdot)$ is both memoryless and strictly monotonically increasing nonlinearity that performs the input-output mapping in order to improve a decision region of the DFE. In the vector form, the relationship between the new input and output is expressed by

$$\mathbf{z}_n = \mathbf{B}\mathbf{x}_n \tag{1}$$

where the input and output vectors and the filter coefficient matrix $\mathbf{B}$ are defined as follows

$$\mathbf{x}_n = \left(x_n, r_{n-1}, ..., r_{n-N}\right)^T,\ \mathbf{z}_n = \left(z_n, r_{n-1}, ..., r_{n-N}\right)^T \tag{2}$$

$$\mathbf{B} = \begin{pmatrix} 1 & b_1 & \cdots & b_N \\ 0 & 1 & \ddots & 0 \\ \vdots & \ddots & \ddots & \ddots \\ 0 & \ddots & \cdots & 1 \end{pmatrix}. \tag{3}$$

Assuming that the previous decisions are correct and the distribution function of the received signal $f_{\mathbf{x}}(x)$ is a constant with respect to coefficients $\{b_j\}$ the following cost function is derived

$$J_{EM} = E\left\{\ln\left|\frac{\partial r_n}{\partial z_n}\right|\right\}. \tag{4}$$

The above cost function is a real scalar function of the coefficients $\{b_j\}$. Its stochastic optimization, in maximizing information, also performs redundancy reduction (i.e., intersymbol interference) [3]. Based on the J_{EM} cost function the CJEM algorithm will be derived for which the inputs, coefficients, activation function and outputs are complex-valued.

[1]Vladimir R. Krstic is with Institute Mihailo Pupin, Volgina 15, 11060 Belgrade, Serbia, E-mail: vladak@kondor.imp.bg.ac.yu

[2]Zoran Petrovic is with the School of Electrical Engineering, University of Belgrade, Bulevar kralja Aleksandra 73, 11000 Belgrade, Serbia, E-mail: zrpetrov@etf.bg.ac.yu

III. COMPLEX ACTIVATION FUNCTION

Let us define the suitable properties of a complex activation function $g(z)=u(z_R,z_I)+iv(z_R,z_I)=r_R+ir_I$, where u and v are its real and imaginary parts and $z=z_R+iz_I$, $i=\sqrt{-1}$; the symbol interval index n has been dropped for convenience of notation. With regard to our intention to find the stochastic gradient of a *real* cost function J_{EM} with respect to complex coefficients, the condition of analyticity, which is mentioned in the introduction, is not so tight. In fact, a complex activation function $g(z)$ need not necessarily be analytic but, it mast have continuous partial derivatives [1], [6]; consequently, it is essential that J_{EM} be *real.* Following this idea, we have assumed that our function $g(z)$ possesses the properties that can be summarized as follows:

P1. $g(z)$ is nonlinear in z_R and z_I.

P2. For all z in a bounded domain D, a suitable complex activation function $g(z)$ must have no singularities (especially no poles) and it must be bounded. This property is, in fact, the bounded-input bounded-output (BIBO) condition for a complex activation function. In other words, the system using the activation function $g(z)$ must be stable in the BIBO sense.

P3. The functions r_R and r_I are continuously differentiable so that the inverses $\delta_R=\left(\frac{\partial r_R}{\partial z_R}\right)^{-1}$ and $\delta_I=\left(\frac{\partial r_I}{\partial z_I}\right)^{-1}$ and the second-order partial derivatives $\frac{\partial^2 r_R}{\partial z_R^2}$, $\frac{\partial^2 r_R}{\partial z_R\partial z_I}$, $\frac{\partial^2 r_I}{\partial z_I^2}$ and $\frac{\partial^2 r_I}{\partial z_I\partial z_R}$ exist and the corresponding quantities $\delta_R\frac{\partial^2 r_R}{\partial z_R^2}$, $\delta_R\frac{\partial^2 r_R}{\partial z_R\partial z_I}$, $\delta_I\frac{\partial^2 r_I}{\partial z_I^2}$ and $\delta_I\frac{\partial^2 r_I}{\partial z_I\partial z_R}$ must be bounded.

P4. The partial derivatives of $g(z)$ obey the condition $\frac{\partial^2 r_R}{\partial z_R^2}\frac{\partial^2 r_I}{\partial z_I^2}\neq\frac{\partial^2 r_I}{\partial z_I\partial z_R}\frac{\partial^2 r_R}{\partial z_R\partial z_I}$. If this condition is not satisfied no learning state is possible for nonzero both input $z_n=(z_R,z_I)$ and $\delta_n=(\delta_R,\delta_I)$. The proof of this proposition is similar with one given in [5] for multilayer perceptron.

IV. COMPLEX JEM ALGORITHM

A. The stochastic gradient of the cost function J_{EM}

The J_{EM} is a function of the absolute value of the complex numbers as well as the cost function J_L of the complex LMS [7] and generalized LMS [1,Ch.17] algorithms. The essential characteristic associating these functions is that they are real scalar functions of the complex coefficients. Based on that fact we can apply the gradient of the generalized LMS algorithm, which is derived for a multilayer perceptron, to find the gradient of JEM.

Let us recall the stochastic gradient of the back-propagation algorithm for the k th neuron in layer $l=M-1$ of multilayer perceptron. For the input-output relation of neuron that is characterized by the nonlinear equation $r=g(z)=u+iv$, $z=z_R+iz_I=\sum_{j=1} w_j x_j$ the gradient with respect to complex coefficients $\{w_j\}$ is defined as follows:

$$\nabla_{w_j}J_L=\frac{\partial J_L}{\partial w_{R,j}}+i\frac{\partial J_L}{\partial w_{I,j}} \tag{5}$$

where

$$\frac{\partial J_L}{\partial w_{R,j}}=\frac{\partial J_L}{\partial u}\left(\frac{\partial u}{\partial z_R}\frac{\partial z_R}{\partial w_{R,j}}+\frac{\partial u}{\partial z_I}\frac{\partial z_I}{\partial w_{R,j}}\right)+\frac{\partial J_L}{\partial v}\left(\frac{\partial v}{\partial z_R}\frac{\partial z_R}{\partial w_{R,j}}+\frac{\partial v}{\partial z_I}\frac{\partial z_I}{\partial w_{R,j}}\right), \tag{6}$$

$$\frac{\partial J_L}{\partial w_{I,j}}=\frac{\partial J_L}{\partial u}\left(\frac{\partial u}{\partial z_R}\frac{\partial z_R}{\partial w_{I,j}}+\frac{\partial u}{\partial z_I}\frac{\partial z_I}{\partial w_{I,j}}\right)+\frac{\partial J_L}{\partial v}\left(\frac{\partial v}{\partial z_R}\frac{\partial z_R}{\partial w_{I,j}}+\frac{\partial v}{\partial z_I}\frac{\partial z_I}{\partial w_{I,j}}\right). \tag{7}$$

By means of the above gradient it is simple to derive the corresponding gradient of J_{EM}. To do that, we have to: i) define the input-output relation of SFBF and ii) observe the variable dependencies in the cost function J_{EM} with the aim to find the corresponding partial gradients using the chain rule of calculus. Following this procedure, both the input and the output of SFBF at the symbol interval n are given by

$$g(z_n)=r_{R,n}+ir_{I,n},$$

$$z_n=(x_{R,n}+ix_{I,n})-\sum_{j=1}^{N}\left(b_{R,n,j}+ib_{I,n,j}\right)\left(r_{R,n-j}+ir_{I,n-j}\right). \tag{8}$$

Consequently, the first-order partial derivatives with respect to the real and imaginary parts of the j th coefficient b_j of the feedback filter are given by

$$\frac{\partial z_{R,n}}{\partial b_{R,n,j}}=r_{R,n-j},\ \frac{\partial z_{R,n}}{\partial b_{I,n,j}}=-r_{I,n-j},$$

$$\frac{\partial z_{I,n}}{\partial b_{R,n,j}}=r_{I,n-j},\ \frac{\partial z_{I,n}}{\partial b_{I,n,j}}=r_{I,n-j}. \tag{9}$$

Using the chain rule of calculus the corresponding second-order partial derivatives are

$$\frac{\partial J_{EM}}{\partial r_R}\frac{\partial r_R}{\partial z_R}=\left(\frac{\partial r_R}{\partial z_R}\right)^{-1}\frac{\partial^2 r_R}{\partial z_R^2},\ \frac{\partial J_{EM}}{\partial r_R}\frac{\partial r_R}{\partial z_I}=\left(\frac{\partial r_R}{\partial z_R}\right)^{-1}\frac{\partial^2 r_R}{\partial z_R \partial z_I},$$
$$\frac{\partial J_{EM}}{\partial r_I}\frac{\partial r_I}{\partial z_I}=\left(\frac{\partial r_I}{\partial z_I}\right)^{-1}\frac{\partial^2 r_I}{\partial z_I^2},\ \frac{\partial J_{EM}}{\partial r_I}\frac{\partial r_I}{\partial z_R}=\left(\frac{\partial r_I}{\partial z_I}\right)^{-1}\frac{\partial^2 r_I}{\partial z_I \partial z_R}. \tag{10}$$

Substituting the Eqs. (9) and (10) in Eqs. (6) and (7) the instantaneous value of gradient of J_{EM} becomes

$$\nabla_{b_j} J_{EM}=\left(\delta_R\frac{\partial^2 r_R}{\partial z_R^2}+i\delta_R\frac{\partial^2 r_R}{\partial z_R \partial z_I}\right)r_{n-j}^* + \left(\delta_I\frac{\partial^2 r_I}{\partial z_I \partial z_R}+i\delta_I\frac{\partial^2 r_I}{\partial z_I^2}\right)r_{n-j}^*,\ j=1,...N. \tag{11}$$

B. The stochastic gradient algorithm of the JEM type

Let us derive the gradient of J_{EM} with respect to the coefficients $\{b_j\}$ for the complex activation function defined by

$$g(z)=r(z)=z(1+\beta|z|^2) \tag{12}$$

where β is a real positive constant. In terms of real and imaginary parts of $r(z)$ the above function is

$$r(z)=z_R\left(1+\beta|z|^2\right)+iz_I\left(1+\beta|z|^2\right). \tag{13}$$

Note that the parameter β in Eq. (13) modifies both the real and the imaginary parts in the same manner. Thus, by means of β we can change the smoothness of the surface (i.e., input-output mapping) that is defined by the magnitude of the complex function $g(z)$.

Now we can prove the properties **P1**-**P4** of the $g(z)$. It is obviously the function is nonlinear in z_R and z_I (**P1**) and bounded in the BIBO sense (**P2**). The normalized second-order partial derivatives in Eq. (11), which are given by

$$\delta_R\frac{\partial^2 r_R}{\partial z_R^2}=\frac{6\beta z_R}{1+\beta|z|^2+2\beta z_R^2},\ \delta_I\frac{\partial^2 r_I}{\partial z_I \partial z_R}=\frac{2\beta z_R}{1+\beta|z|^2+2\beta z_I^2},$$
$$\delta_R\frac{\partial^2 r_R}{\partial z_R \partial z_I}=\frac{2\beta z_I}{1+\beta|z|^2+2\beta z_R^2},\ \delta_I\frac{\partial^2 r_I}{\partial z_I^2}=\frac{6\beta z_I}{1+\beta|z|^2+2\beta z_I^2}, \tag{14}$$

are bounded. It is clear that they must be bounded since coefficients updating is in quantities proportional to the normalized partial derivatives (**P3**). Finally, second-order partial derivatives of $g(z)$ are given by

$$\frac{\partial^2 r_R}{\partial z_R^2}=6\beta z_R,\ \frac{\partial^2 r_R}{\partial z_R \partial z_I}=2\beta z_I,$$
$$\frac{\partial^2 r_I}{\partial z_I^2}=6\beta z_I,\ \frac{\partial^2 r_I}{\partial z_I \partial z_R}=2\beta z_R. \tag{15}$$

Thus, $36\beta^2 z_R z_I \neq 4\beta^2 z_R z_I$ and **P4** is satisfied.

The stochastic gradient of J_{EM} in Eq. (11) is function of the normalized second-order partial derivatives that are given by relations in Eq. (14) and that can be approximated with their Taylor series expansions. Using the approximation method $\frac{1}{1+x}=1-x+x^2-...\approx 1-x$ and some simple algebra the gradient given by Eq. (11) becomes

$$\nabla_{b_j} J_{EM}=\left[8\beta z_R-8\beta^2 z_R|z|^2-4\beta^2 z_R z_I^2-12\beta^2 z_R^3\right]r_{n-j}^* + i\left[8\beta z_I-8\beta^2 z_I|z|^2-4\beta^2 z_I z_R^2-12\beta^2 z_I^3\right]r_{n-j}^*,\ j=1,...,N. \tag{16}$$

The above gradient can be further simplified dropping the terms $4\beta^2 z_R z_I^2$, $12\beta^2 z_R^3$, $4\beta^2 z_I z_R^2$ and $12\beta^2 z_I^3$. This approximation is motivated by the aim to provide the low computation complexity CJEM algorithm comparable with algorithms of the Bussgang type. Thus, the gradient of J_{EM} is given by

$$\nabla_{b_j} J_{EM}=8\beta\left[z\left(1-\beta|z|^2\right)\right]r_{n-j}^*,\ j=1,...,N. \tag{17}$$

and the corresponding CJEM stochastic gradient algorithm is

$$b_{j,n+1}=b_{j,n}+\mu\nabla_{b_j} J_{EM},$$
$$b_{j,n+1}=b_{j,n}+\mu_S z_n\left(1-\beta|z_n|^2\right)r_{n-j}^*,\ j=1,...,N, \tag{18}$$

where $\mu_S=8\mu\beta$ is a step size. It should be noted that the smoothing parameter β could be used as a tool to vary the convergence characteristics of the JEM algorithm. In other words, varying the parameter β it is possible to optimize the input-output mapping in a sense of maximum joint entropy.

V. Simulation Results

In this section we present the convergence characteristics of the "self-adaptive" DFE (Soft-DFE) that combines both the adaptive structure and the different adaptation criteria [8]. In this scheme, the CJEM algorithm provides a soft transition from blind to tracking operation mode.

The self-adaptive Soft-DFE carries out a difficult task of blind activation as a sequence of several subtasks progressing from easier to more complex, as illustrated in Fig 2. In blind

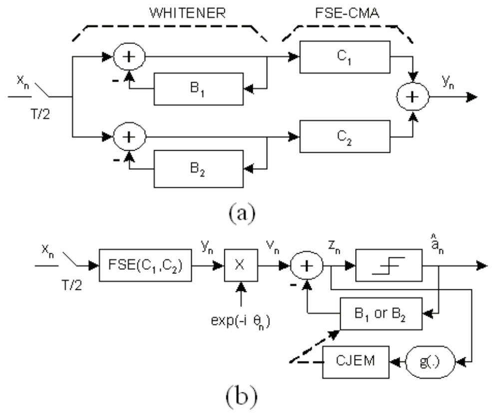

Fig. 2. The structure of Soft-DFE: (a) linear T/2 fractionally-spaced equalizer (FSE) in blind acquisition mode represented by two symbol-rate equalizers, (b) decision-directed DFE with modified CJEM soft feedback filter in transition mode

acquisition mode, the Soft-DFE is the linear equalizer consisting of: (i) the SFBF-FSE cascade where the SFBF is a pure recursive whitener that compensates, in predictive manner, the amplitude distortion of perhaps nonminimum phase channel and (ii) the FSE that compensates the phase distortion of channel-whitener combination. In tracking mode, the Soft-DFE is a classical DFE performing MMSE decision-directed LMS equalization. Besides, the Soft-DFE is characterized by the soft transition mode that takes place immediately after structure switching, Fig 2b. In this phase of operation, the objective of the SFBF is to mitigate the error propagation effects that emerge at the moment of both the structure and criteria switching.

The length of Soft-DFE is $22T$ and $6T$ in its FSE and SFBF parts, respectively, where T denotes symbol interval. The initial values of coefficients are all zero except for two centered reference taps in FSE. The channel impulse response, comprising the pulse-shaping filter at transmitter and the multipath propagation channel, is given by

$$h(t) = e(t)W(t) + a_1 e(t-\tau_1)W(t-\tau_1) + a_2 e(t-\tau_2)W(t-\tau_2) \qquad (19)$$

where $e(t)$ is the basic pulse and $W(t)$ is a rectangular window spanning $\{-16T, 16T\}$. The presented simulation results are carried out through use a channel with the following propagation parameters: $a_1 = 0.9$, $a_2 = 0.4$, $\tau_1 = 3T/4$ and $\tau_2 = 2T$.

The convergence characteristics of the Soft-DFE are evaluated running Monte Carlo tests with 1000 independent runs. The results of testing include the MSE convergence characteristics presented in Fig. 3 where threshold levels $M_{TL1} = 1.5dB$ and $M_{TL2} = -8dB$ define the soft transition mode. The threshold level M_{TL1} is selected to be a high enough so that the constellation-eye of 16QAM signal is practically closed. The presented results should highlight the effect of the smoothing parameter β, which is selected in the range $\beta = \{1,2,4,6,8,10\}$, on the Soft-DFE convergence. For

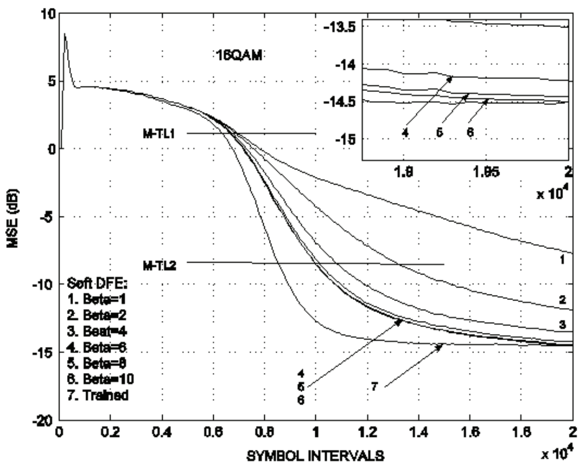

Fig. 3. MSE convergence of Soft-DFE for 16QAM and SNR=25 dB: CJEM for different values of smoothing parameter $\beta = \{1,2,4,6,8,10\}$ and trained Soft-DFE

the purpose of comparison, the Soft-DFE with the CJEM is compared with the Soft-DFE that, after blind acquisition, switches into the trained mode based on the desired symbols a_n and LMS algorithm. This scheme is denoted as the "trained" Soft-DFE. Obviously, by increasing β, the CJEM algorithm provides a promising MSE convergence. It can be estimated that Soft-DFE strikes the best MSE convergence for β in range $\{6,10\}$. In the contrary, for smaller values of β, CJEM algorithm loses power so that the convergence characteristics become unacceptable.

The presented results can be summarized as follows: (i) the proposed complex nonlinearity has a property to maximize joint entropy, (ii) the simulation results indicate that the performance of CJEM algorithm can be optimized, for given signal constellation, by means of the smoothing parameter.

REFERENCES

[1] S. Hykin, *Adaptive Filter Theory*, Forth edition, Prentice-Hall 2002.

[2] Y.H. Kim, S. Shamsunder, "Adaptive Algorithms for Channel Equalization with Soft Decision Feedback," *IEEE JSAC*. vol. 16, Dec. 1998.

[3] A.J. Bell, T.J. Sejnowski, "An information-maximization approach to blind separation and blind deconvolution," *Neural Computition,* vol 7, pp. 1129-1159, 1995.

[4] N. Benvenuto, F. Piazza, "On the Complex Backpropagation Algorithm," *IEEE Trans. Signal Processing*, pp.967-969, Apr. 1992.

[5] G.M. Georgiou, C. Koutsougeras, "Complex Domain Backpropagation," *IEEE Trans. Circuits and Systems*, pp.330-334, May 1992.

[6] C. You, D.Hong, "Nonlinear Blind Equalization Schemes Using Complex-Valued Multilayer Feedforward Neural Networks," *IEEE Trans. Neural Networks*, pp.1442-1455, Nov, 1998.

[7] B. Widrow, J. McCool, M. Ball, "The complex LMS algorithm," *Proc. IEEE,* vol.63, pp. 719-720, Apr. 1975.

[8] V.R. Krstic, Z. Petrovic, "Decision Feedback Blind Equalizer with Maximum Entropy," EUROCON 2005, Belgrade, November , 2005