Complex Domain Decision Feedback Equalizer Based on Bell-Sejnowski Neuron

V. R. Krstic, *Member, IEEE*

Abstract **– This paper considers the complex domain Bell-Sejnowski class neuron in the self-adaptive decision feedback equalizer (DFE) structure where the joint entropy is maximized through an appropriate complex loopback nonlinearity. The presented simulation results are particularly focused on the transition phase when DFE switches configuration in order to gain a deeper insight into the ability of the new joint entropy maximization algorithm to mitigate error propagation effects.**

Keywords **– Blind decision feedback equalization, Bell-Sejnowski class neuron model**

I. Introduction

ONE of the most interesting aspects of the independent component analysis theory proposed by Bell and Sejnowski [1] is the ability of the neurons in the structural model to align to the statistical distributions of the input signals. Such observation was successfully exploited in order to design different learning rules related to the broad range of signal processing applications such as blind separations, blind deconvolution and probability density function estimation [2]. More particularly in the field of blind equalization methods, the basic Bell-Sejnowski maximum-entropy adapting formula was applied to the decision feedback equalizer (DFE), i.e., its decision feedback filter and the class of joint entropy maximization (JEM) stochastic gradient algorithms was derived for the real-valued inputs and coefficients [3]. Also, more recently in [4]-[6], this JEM method has been extended to the complex domain in the self-adaptive DFE structure [7].

In this paper we consider the soft feedback (SFBF) filter of the Bell-Sejnowski class in the complex domain referring to the complex stochastic gradient algorithm of the JEM type (CJEM) that has been derived in [6] for the specific complex activation function. Using a heuristic approach this algorithm has also been extended to an all-pole whitener which is a component of the self-adaptive DFE (Soft-DFE) [4], [5]. In particular, our main goal in this paper is to gain a deeper insight into the relationship between the smoothness of the activation function and the error propagation phenomenon. Using the software simulations it is shown that the error propagation effects can be mitigated by means of the smoothing parameter of the new CJEM type algorithms.

V. R. Krstic, Institute Mihailo Pupin, Belgrade, Serbia (tel: 381-11-2772953; fax: 381-11-2755 978; e-mail: vladak@kondor. imp.bg.ac.yu).

II. Complex Domain Decision Feedback Equalizer Maximizing Joint Entropy

A. DFE with neuron of Bell-Sejnowski class

The SFBF filter in [3] was considered as a neural unit with the corresponding set of adjustable parameters, Fig 1. In this context, the input-output description of an activation function $g(\cdot)$, which is both a memoryless and a strictly monotonically increasing nonlinearity, is given as follows: the neuron's net input z_n is the combination of the noiseless channel output x_n mixing present and past data symbols and the weighted sum of previous outputs r_{n-j}, $j=1,...,N$, where N is filter length. The objective of the activation function in the given SFBF structure model is to transform the input sequence z_n with an unknown probability density function to a maximum entropy sequence $r_n = g(z_n)$ using the Bell-Sejnowski type adapting formula. In other words, the coefficients $\{b_j\}$ are the subject of learning through an adaptation algorithm that maximizes the joint entropy defined by $H\left[r_{n,1},...,r_{n,N+1}\right] \triangleq -E\{\ln f_r(\mathbf{r}_n)\}$ where $f_r(\mathbf{r}_n)$ is the joint density function of the input vector $\mathbf{r}_n = \left[r_{n,1},...,r_{n,N+1}\right]^T$.

Based on assumptions commonly used for DFE analysis, that is, the input data a_n are independently identically distributed and zero-mean and the previous decisions are correct, $r_{n-j} = a_{n-j}$, the JEM criterion was suitably simplified as $\widetilde{J}_{EM} = E\{\ln|J|\} = E\left\{\ln\left|\frac{\partial r_n}{\partial z_n}\right|\right\}$ where E is expectation and $|J| = \left|\frac{\partial r_n}{\partial z_n}\right|$ is the absolute

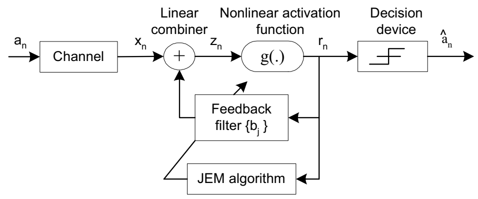

Fig. 1 Model of transmission system with the basic scheme of soft feedback filter

value of the Jacobian of the transformation. Consequently, in the case of stochastic gradient ascent learning the corresponding instantaneous approximation of $\tilde{J}_{EM}$ can be used as the JEM criterion, that is,

$$J_{EM} = \ln\left|\frac{\partial r_n}{\partial z_n}\right|. \quad (1)$$

The above entropy cost function expresses the main idea underlying the use of information-theoretic concept in channel deconvolution applications: the statistical dependence between the current output of the SFBF and its previous outputs can be reduced by maximizing the entropy J_{EM} in an iterative manner and, hence, that leads to reduction of intersymbol interference (ISI).

B. Complex domain gradient of cost function J_{EM}

Let the neuron's complex-valued activation function be $g(z) = r_R(z_R, z_I) + ir_I(z_R, z_I)$ where the subscripts R and I indicate the real and imaginary parts of g, respectively, and $z = z_R + iz_I$, $i = \sqrt{-1}$; for convenience of notation, the symbol interval index n has been dropped. The function $g(z)$ is assumed to be an appropriate nonlinearity whose properties will be discussed later. In the complex-valued case of interest here, the input-output relationship of the SFBF can be defined by the net $z = z_R + iz_I = (x_R + ix_I) - \sum_{j=1}^{N}\left(b_{j,R} + ib_{j,I}\right)\left(r_{j,R} + ir_{j,I}\right)$. It should be noted that for the given complex net the entropy J_{EM} is a real scalar function of unknown complex coefficients. In fact, this is an essential property of function J_{EM} that simplifies deriving of its gradient with respect to complex coefficients. This simplification is done through use of a multidimensional complex gradient vector [8, Ch. 2], whose *j*th element is defined as follows

$$\nabla_{b_j} J_{EM} = \frac{\partial J_{EM}}{\partial b_{R,j}} + i\frac{\partial J_{EM}}{\partial b_{I,j}}, \; j = 1,...N. \quad (2)$$

In our particular case of the SFBF's net, the gradient of J_{EM} with respect to coefficients $\{b_j\}$, which is derived in [6], is given by

$$\nabla_{b_j} J_{EM} = -\left[\left(r_R^{rr} + ir_R^{ri}\right)\delta_R + \left(r_I^{ir} + ir_I^{ii}\right)\delta_I\right]r_{n-j}^{*}, \quad (3)$$

where * denotes complex conjugation. The respective quantities in (3) are the corresponding first and second order partial derivatives of the activation function which are defined as follows: $\left(\frac{\partial r_R}{\partial z_R}\right)^{-1} = \delta_R$, $\left(\frac{\partial r_I}{\partial z_I}\right)^{-1} = \delta_I$, $\frac{\partial^2 r_R}{\partial z_R^2} = r_R^{rr}$, $\frac{\partial^2 r_R}{\partial z_R \partial z_I} = r_R^{ri}$, $\frac{\partial^2 r_I}{\partial z_I^2} = r_I^{ii}$ and $\frac{\partial^2 r_I}{\partial z_I \partial z_R} = r_I^{ir}$. It is worth noting in (3) that the entropy gradient does not depend explicitly on an activation function, but it is governed by the quantities that are given as proportions of the corresponding second and first order partial derivatives of a complex activation function.

Let us discuss in more detail the properties of the complex function $g(z)$. This function need not necessarily be analytic [8, Ch. 17], but it is the suitable activation function in the sense of properties that provide a stable gradient learning [6] and 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 second-order partial derivatives r_R^{rr}, r_R^{ri}, r_I^{ii}, r_I^{ir} exist for all $z \in \mathbb{C}$, and the corresponding "normalized" derivatives $\delta_R r_I^{rr}$, $\delta_R r_R^{ri}$, $\delta_I r_I^{ii}$ and $\delta_I r_I^{ir}$ must be bounded since coefficients updating is in quantities proportional to the normalized second-order partial derivatives (see (3)). Obviously, the derivatives δ_R and δ_I must be different from zero to avoid an instability.

P4. The partial derivatives of $g(z)$ obey $r_R^{rr} r_I^{ii} \neq r_I^{ir} r_R^{ri}$. If this condition is not satisfied no learning state is possible for both nonzero input $z_n = (z_R, z_I)$ and $\delta_n = (\delta_R, \delta_I)$.

C. Stochastic gradient algorithms of the JEM type

Let us consider a simple complex activation function which is described by the input-output equation

$$g(z) = r(z) = z(1 + \beta|z|^2) \quad (4)$$

where β is a real positive constant. This function has the property of mapping a point $z = z_R + iz_I = (z_R, z_I)$ on the complex plane to a unique point $g(z) = \left(z_R\left(1 + \beta|z|^2\right)\right), \left(z_I\left(1 + \beta|z|^2\right)\right)$ keeping the same phase angle. In addition, the magnitude of the complex function $g(z)$ is a paraboloid that has the following properties: the horizontal cross sections are actually circles, the bottom surface is located at the point $(0,0)$ and the parameter β modifies the shape of the surface. In fact, the parameter β affects the input-output mapping by modifying, in the same manner, both the real and the imaginary parts of activation function. Finally, it can be easily proved that this function is not analytic, but it satisfies the properties **P1-P4**.

For the complex function defined by (4) the stochastic gradient algorithm of the JEM type (CJEM) is given by

$$b_{j,n+1} = b_{j,n} + \alpha \nabla_{b_{n,j}} J_{EM} ,$$

$$b_{j,n+1} = b_{j,n} - \mu z_n \left(1 - \beta \left|z_n\right|^2\right) r_{n-j}^* , \; j = 1,...,N , \qquad (5)$$

where α is a positive learning constant and $\mu = 8\alpha\beta$ is a step size. It is worth noting that the parameter β, which controls the shape of the paraboloid, can be used as a tool to optimize the convergence characteristics of the CJEM algorithm. We have denoted β as a smoothing parameter taking into account its overall affects to algorithm behavior in the structure model.

The CJEM algorithm is a key component of the Soft-DFE system [4], [5] that is based on the fractionally-spaced structure illustrated in Fig. 2. In this self-adaptive structure the filters (B_1, B_2) start their adaptation as all-pole whiteners in the blind acquisition mode and then, one of them, continues adaptation as a part of the SFBF in transition and tracking modes [4]. In this scenario, the two variants of the CJEM algorithm are of interest. In the first case, by the analogy with the Extended LMS algorithm [7], the algorithm (5) can be transformed into the corresponding decorrelation algorithm

$$b_{j,n+1} = b_{j,n} - \mu_E u_n \left(1 - \beta_E \left|u_n\right|^2\right) u_{n-j}^* , \; j = 1,...,N \qquad (6)$$

where μ_E is a step size and β_E is a parameter that can be suitably selected to provide the fast and stable convergence of the whiteners. In [5], the BIBO stability of the whiteners has been proved using the scheme where the gain control (GC) is coupled with the output of one of whiteners.

The second variant of the CJEM algorithm is the result of the structure modification in Fig. 1 where now the hard decision estimates of the transmitted symbols, $\hat{a}_{n-j}$, feed the feedback filter instead of the soft decisions r_{n-j} (see Fig. 2b). Consequently, the decision-directed variant of the CJEM algorithm (DD-CJEM) reads

$$b_{j,n+1} = b_{j,n} - \mu z_n \left(1 - \beta \left|z_n\right|^2\right) \hat{a}_{n-j}^* , \; j = 1,...,N . \qquad (7)$$

The applied modification is a heuristic approach that is motivated by the following reasons: (i) the whitener is possibly correctly set up during the blind acquisition mode because it is an amplitude equalizer performing a speed convergence even if a channel is in a deep fade, (ii) the fractionally-spaced equalizer (FSE) of sufficient length that is controlled by the constant modulus algorithm (FSE-CMA) is globally convergent [9] and (iii) the DD-CJEM

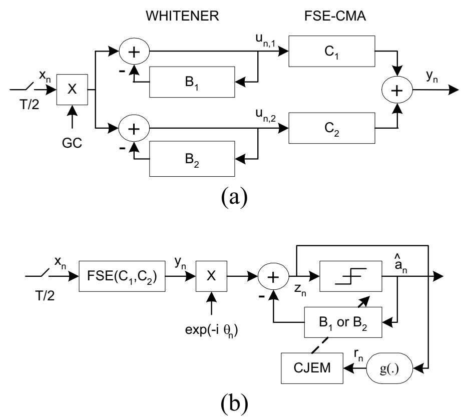

Fig. 2. The structure of Soft-DFE:
(a) linear T/2 fractionally spaced equalizer (FSE-CMA) in blind acquisition mode, (b) decision-directed DFE with modified CJEM soft feedback filter in transition mode

algorithm combines the soft error $z_n \left(1 - \beta \left|z_n\right|^2\right)$ of the CJEM algorithm and the conventional decision-directed method.

III. Simulation Results

The Soft-DFE achievements presented in this paper are characterized in terms of symbol error rate (SER) and block error length (BEL) statistics which are observed in the soft transition mode. We have focused on the transition mode supposing that the blind acquisition is successfully accomplished by using the decorrelation algorithm (6) for $\beta_E = 1.0$. The presented results are obtained using the 16QAM signal, multipath stationary channels (MP) and equalizers length of $22T$ and $6T$ (T is symbol interval) in their linear and recursive parts, respectively. The multipath environment is realized by the three-ray channel model included in the transmitter filter, the amplitude characteristics of which are presented in Fig. 3a for different propagation parameters (attenuation and propagation delay).

Let us first describe the transition mode of the Soft-DFE by using its convergence characteristics obtained with the Mp channels and the signal-to-noise ratio $SNR = 25.0dB$, Fig. 3b. The mean-square error (MSE) threshold levels $M_{TL-1} = 1.5dB$ and $M_{TL-2} = -8.0dB$ define the beginning and the end of the transition mode, respectively. The threshold level M_{TL-1} is selected to be sufficiently high so that, at the moment of configuration switching, the initial burst of errors is certainly present. In fact, this is the most critical phase of the Soft-DFE operation when several algorithms are directed and slightly coupled by the sequence of unreliable symbol estimates: (i) the DD-LMS algorithm that adjusts T/2 FSE, (ii) phase-tracking loop that rotates a signal constellation and (iii) the DD-CJEM algorithm.

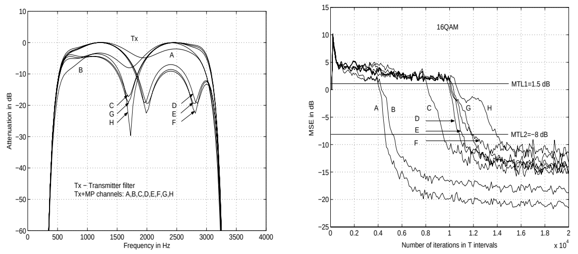

Fig. 3 (a) Amplitude characteristics of MP channels (A,B,C,D,E,F,G,H), (b) MSE convergence characteristics of the Soft-DFE with MP channels for $\beta_E = 1$ and $\beta = 10$

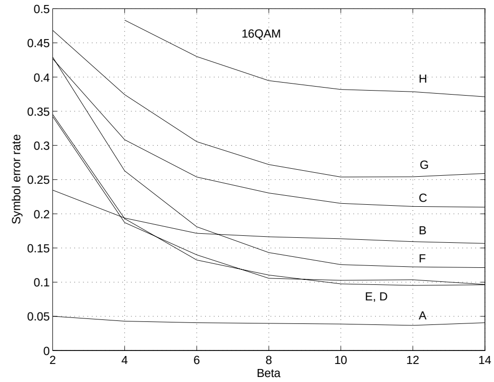

Fig. 4 Symbol error rate versus β in transition mode: Mp=(A,B,C,D,E,F,G,H), 16QAM, SNR=25 dB

Fig. 4 presents the results of the SER versus the smoothing parameter β which are averaged over 200 independent Monte Carlo runs. The SER is measured during 1500 symbol intervals that, in the case of severe-ISI channels, Mp=(C,D,E,F,G,H), roughly corresponds to equalizer's convergence time between thresholds M_{TL-1} and M_{TL-2}. Obviously, for these channels, the SER decreases with increasing β in a similar manner. This behavior clearly indicates the capability of DD-CJEM to mitigate error propagation to some extent. In Table 1, the block error length statistics for the lengths from 2 to 8 in T intervals are shown; the BEL=8 also includes longer burst errors. The long block errors are disappeared and the number of short blocks is reduced for β in range $\{10,14\}$. It should be noted that the SER and BEL statistics, which are introduced as a measure of the error propagation phenomenon, are not critically sensitive for a wide range of values of the smoothing parameter.

The presented simulation results have proved the ability of the new algorithm to mitigate the error propagation effects. Also, these results indicate that improved bust error statistics (no long burst errors) can be transform into larger channel coding gain in systems with separate equalization and decoding scheme [10].

TABLE 1: BLOCK ERROR LENGTH STATISTICS FOR MP-E

BEL/β	2	3	4	5	6	7	8
2	71	29	13	7	3	1	2
4	40	15	6	3	1	1	1
6	27	10	4	2	1	0	0
8	23	8	3	2	0	0	0
10	21	7	2	1	0	0	0
12	21	6	2	1	0	0	0
14	21	6	2	1	0	0	0

IV. CONCLUSIONS

The purpose of this contribution is to discuss in some detail the behavior of the self-adaptive DFE based on the complex-valued neuron of the Bell-Sejnowski class in order to find a relationship between the smoothing parameter of the chosen activation function and the error propagation effects. The presented simulation results have proved the ability of the new algorithm to mitigate the error propagation effects for a wide range of values of the smoothing parameter. It is possible to determine the range of values of β that provides the best performance of the algorithm for a known signal constellation. This is of the practical importance because the overall performance characteristics of the Soft-DFE can be optimized selecting β in the given range. Hence, the DD-CJEM algorithm becomes a nonparametric stochastic gradient algorithm with the same low computation complexity as the CMA.

REFERENCES

[1] A.J.Bell, T.J.Sejnowski, "An information maximization approach to blind separation and blind deconvolution," *Neural Computation,* vol 7, 1996, pp. 1129-1159.

[2] S. Fiori, "Notes on Bell-Sejnowski PDF-Matching Neuron," *Neural Computation* Vol.14, No.12, Dec. 2002, pp. 2847-2855.

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

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

[5] V. R. Krstic, Z. Petrovic, "Blind Decision Feedback Equalizer: Stability of the Decorrelator with Maximum Entropy," 50th Conference for Electronics, Telecommunications, Computers, Automation, and Nuclear Engineering, Belgrade, June 6-8, 2006.

[6] V. R. Krstic, Z. Petrovic, "Complex-Valued Maximum Joint Entropy Algorithm for Blind Decision Feedback Equalizer, "8th International Conference on Telecommunications in Modern Satellite, Cable and Broadcasting Services-TELSIKS 2007, Serbia, Niš, September 26-28, 2007, pp. 601-604.

[7] Labat, J., Macchi, O., and Laot, C.: "Adaptive Decision Feedback Equalization: Can You Skip the Training Period?," *IEEE Trans. Commun..,* July 1998, pp.921-930.

[8] Hykin, S.: "Adaptive Filter Theory" (Prentice-Hall 2002 4th edn.)

[9] Johnson, C.R.Jr. et al. : "The core of FSE-CMA Behavior Theory". In S.Haykin (Ed.), *Unsupervised adaptive filtering: Vol II. Blind deconvolution* (pp. 13-112). New York: John Wiley & Sons, 2000.

[10] J.-T. Liu, S.B.Gelfand, "Optimized Decision-Feedback Equalization for Convolutional Coding With Reduced Delay," IEEE Trans. Commun., Nov., 2005, pp.1859-1866.