# COMPLEX BACKPROPAGATION NEURAL NETWORK USING ELEMENTARY TRANSCENDENTAL ACTIVATION FUNCTIONS

*Taehwan Kim*[†,*] *and Tülay Adali*[*]

[†]The MITRE Corporation
McLean, Virginia 22102, U.S.A.

[*]Department of Computer Science and Electrical Engineering
University of Maryland Baltimore County
Baltimore, Maryland 21250, U.S.A.

## ABSTRACT

Designing a neural network (NN) for processing complex signals is a challenging task due to the lack of bounded and differentiable nonlinear activation functions in the entire complex domain **C**. To avoid this difficulty, *'splitting'*, i.e., using uncoupled real sigmoidal functions for the real and imaginary components has been the traditional approach, and a number of fully complex activation functions introduced can only correct for magnitude distortion but can not handle phase distortion. We have recently introduced a fully complex NN that uses a hyperbolic tangent function defined in the entire complex domain and showed that for most practical signal processing problems, it is sufficient to have an activation function that is bounded and differentiable *almost everywhere* in the complex domain. In this paper, the fully complex NN design is extended to employ other complex activation functions of the hyperbolic, circular, and their inverse function family. They are shown to successfully restore nonlinear amplitude and phase distortions of non-constant modulus modulated signals.

## 1. INTRODUCTION

The main reason for the difficulty in finding a nonlinear complex activation function in NN design is the conflict between the boundedness and the differentiability of complex functions in the entire complex plane, as stated in Louiville's theorem [1]. It states that *a bounded entire function must be a constant in* **C**, where an entire function is defined as analytic, i.e., differentiable at every point in **C** [1]. To combat this unbounded nature of analytic functions in **C**, two fully complex activation functions, proposed by Georgiou and Koutsougeras [2], and Hirose [3] have normalized or scaled the amplitude of complex signals. However, these functions preserve the phase, thus are incapable of learning the phase variations between the input and the target in static pattern matching applications where the input layer does not include time delay elements. Even in time-delayed NN structure, they perform poorly in restoring nonlinear amplitude and phase distortion of non-constant modulus signals due to their radial mapping characteristics as shown in the numerical examples in Section 4.

In contrast, another approach to process complex signals by NNs is to use 'split' complex activation with two real-valued activation functions for the in-phase (I) and quadrature (Q) components [4]-[7]. While this approach could avoid the unboundedness of fully complex activation functions in view of Louiville's theorem, the split complex activation function could never be analytic.

The split approach typically employs a pair of tanh $x$, $x \in$ **R,** functions as the complex activation function in a feed-forward NN (FNN) structure [4]-[6]. We have recently shown that complex hyperbolic tangent function tanh $z$ can be successfully used as a fully complex activation function. Even though tanh $z$ is periodic and has periodic singularities, it still outperforms conventional split complex backpropagation and least mean squares (LMS) algorithms in nonlinear channel equalization of quadrature phase shift keying (QPSK) modulated signals when the domain is bounded in the neighborhood of the unit circle. This is possible because the bounded and well-defined derivatives meet the Cauchy-Riemann equations almost everywhere in **C** while not limited by the Louiville's theorem in the unit circle neighborhood.

In this paper, the fully complex NN is extended to employ eight other elementary transcendental functions. It is shown that Cauchy-Riemann equations can be used to simplify the fully complex backpropagation algorithm derived in [2]. They also help to relax the properties a fully complex activation function is 'desired' to possess as defined in [2] and [7]. The performance of the complex backpropagation algorithm using these complex activation functions is compared with split complex backpropagation and complex LMS (CLMS) schemes in numerical examples. The Volterra series nonlinear satellite Traveling Wave Tube Amplifier (TWTA) model exhibiting amplitude-to-amplitude (AM/AM) and amplitude-to-phase (AM/PM) distortion is used as the nonlinear channel model. Using this model, since several backpropagation algorithms showed similar performances, not only unsaturated but also saturated amplification condition was tested for the severe nonlinear distortion of non-constant modulus signal being used as input to the supervised learning process.

## 2. COMPLEX BACKPROPAGATION

Cauchy-Riemann equations can be used to simplify the fully complex backpropagation algorithm derived in [2] as shown next. Cauchy-Riemann equations are the necessary condition for a complex function to be analytic at a point $z \in \mathbf{C}$ and can be written by noting that the partial derivatives of $f(z) = u(x,y) + iv(x,y)$, where $z=x+iy$, should be equal along the real and imaginary axes:

$$f'(z) = u_x + iv_x = v_y - iu_y \tag{1}$$

Equating the real and imaginary parts in (1), we obtain the *Cauchy-Riemann* equations: $u_x = v_y, v_x = -u_y$. Also note that this enables equation (1) to be expressed more concisely as

$$f'(z) = f_x = -if_y . \tag{2}$$

For the fully complex activation function $f(z)$, the squared error at the output layer is written as

$$E = \frac{1}{2}\sum_n |e_n|^2, \quad e_n = d_n - o_n \tag{3}$$

$$o_n \equiv f(z_n) \equiv u_n + iv_n,\ z_n \equiv x_n + iy_n \equiv \sum_k W_{nk} X_{nk} = \sum_k (W_{nkR} + iW_{nkI})(X_{nkR} + iX_{nkI}) \tag{4}$$

where $d_n$ is the $n$-th desired symbol and $o_n$ is the output of the $n$-th output neuron and the subscripts $R$ and $I$ indicate the real and imaginary components, respectively. The backpropagation weight adaptation rule requires the computation of the gradient $\partial E/\partial W_{nk}$. The gradient of the error function with respect to the real and imaginary components of $W_{nk}$ can be written as

$$\partial E/\partial W_{nk} \equiv \nabla_{W_{nk}} E = \partial E/\partial W_{nkR} + i\,\partial E/\partial W_{nkI} \tag{5}$$

and using the chain rule,

$$\frac{\partial E}{\partial W_{nkR}} = \frac{\partial E}{\partial u_n}\left(\frac{\partial u_n}{\partial x_n}\frac{\partial x_n}{\partial W_{nkR}} + \frac{\partial u_n}{\partial y_n}\frac{\partial y_n}{\partial W_{nkR}}\right) + \frac{\partial E}{\partial v_n}\left(\frac{\partial v_n}{\partial x_n}\frac{\partial x_n}{\partial W_{nkR}} + \frac{\partial v_n}{\partial y_n}\frac{\partial y_n}{\partial W_{nkR}}\right) \tag{6}$$

$$\frac{\partial E}{\partial W_{nkI}} = \frac{\partial E}{\partial u_n}\left(\frac{\partial u_n}{\partial x_n}\frac{\partial x_n}{\partial W_{nkI}} + \frac{\partial u_n}{\partial y_n}\frac{\partial y_n}{\partial W_{nkI}}\right) + \frac{\partial E}{\partial v_n}\left(\frac{\partial v_n}{\partial x_n}\frac{\partial x_n}{\partial W_{nkI}} + \frac{\partial v_n}{\partial y_n}\frac{\partial y_n}{\partial W_{nkI}}\right) \tag{7}$$

Defining $\delta_n \equiv -\partial E/\partial u_n - i\,\partial E/\partial v_n$, $\delta_{nR} \equiv -\partial E/\partial u_n$ and $\delta_{nI} \equiv -\partial E/\partial v_n$, and using the following partial derivatives identifiable from (4),

$$\frac{\partial x_n}{\partial W_{nkR}} = X_{nkR}, \frac{\partial y_n}{\partial W_{nkR}} = X_{nkI}, \frac{\partial x_n}{\partial W_{jkI}} = -X_{nkI}, \frac{\partial y_n}{\partial W_{nkI}} = X_{nkR}$$

(6) and (7) can be simplified as

$$\frac{\partial E}{\partial W_{nkR}} = -\delta_{nR}\left(\frac{\partial u_n}{\partial x}X_{nkR} + \frac{\partial u_n}{\partial y}X_{nkI}\right) - \delta_{nI}\left(\frac{\partial v_n}{\partial x}X_{nkR} + \frac{\partial v_n}{\partial y}X_{nkI}\right) \tag{8}$$

$$\frac{\partial E}{\partial W_{nkI}} = -\delta_{nR}\left(\frac{\partial u_n}{\partial x}(-X_{nkI}) + \frac{\partial u_n}{\partial y}X_{nkR}\right) - \delta_{nI}\left(\frac{\partial v_n}{\partial x}(-X_{nkI}) + \frac{\partial v_n}{\partial y}X_{nkR}\right). \tag{9}$$

Combining (8) and (9), the gradient of the error function becomes

$$\nabla_{W_{nk}} E = -\overline{X}_{nk}\left\{\left(\frac{\partial u_n}{\partial x} + i\frac{\partial u_n}{\partial y}\right)\delta_{nR} + \left(\frac{\partial v_n}{\partial x} + i\frac{\partial v_n}{\partial y}\right)\delta_{nI}\right\} \tag{10}$$

Note that (10) is in the same form given in [2], but further applying the *Cauchy-Riemann* equations, a more compact representation for the gradient of error function is obtained using the simple derivative form given in (2)

$$\nabla_{W_{nk}} E = -\overline{X}_{nk}\cdot\left(\frac{\partial \bar{f}}{\partial x}\delta_{nR} - i\left(-i\frac{\partial \bar{f}}{\partial y}\right)\delta_{nI}\right) = -\overline{X}_{nk}\,\overline{f}'(z)\overline{\delta}_n. \tag{11}$$

It is worth noting that except the conjugate on each term, it is identical to the gradient of the error function in the real version of the backpropagation algorithm as would be expected.

Complex weight update $\Delta W_{nk}$ is proportional to the negative gradient:

$$\Delta W_{nk} = \alpha \overline{X}_{nk}\,\overline{f}'(z)\overline{\delta}_n \tag{12}$$

where $\alpha$ is a real positive learning rate. When the complex weight belongs to an output neuron:

$$\delta_n = e_n = d_n - o_n \tag{13}$$

and when $W_{mk}$ belongs to the $m$-th hidden layer, the net input $z_m$ to neuron $m$ is the same as in equation (4):

$$z_m = x_m + iy_m = \sum_k (u_k + iv_k)(W_{mkR} + iW_{mkI}),$$

where the index $k$ belongs to every neuron that feeds into the neuron $m$. Using the chain rule,

$$\delta_{mR} = -\frac{\partial E}{\partial u_m} = -\sum_k \frac{\partial E}{\partial u_k}\left(\frac{\partial u_k}{\partial x_k}\frac{\partial x_k}{\partial u_m} + \frac{\partial u_k}{\partial y_k}\frac{\partial y_k}{\partial u_m}\right) - \sum_k \frac{\partial E}{\partial v_k}\left(\frac{\partial v_k}{\partial x_k}\frac{\partial x_k}{\partial u_m} + \frac{\partial v_k}{\partial y_k}\frac{\partial y_k}{\partial u_m}\right)$$

$$= \sum_k \delta_{kR}\left(\frac{\partial u_m}{\partial x_k}W_{mkR} + \frac{\partial u_m}{\partial y_k}W_{mkI}\right) + \sum_k \delta_{kI}\left(\frac{\partial v_m}{\partial x_k}W_{mkR} + \frac{\partial v_m}{\partial y_k}W_{mkI}\right) \tag{14}$$

Similarly,

$$\delta_{mI} = -\frac{\partial E}{\partial v_m} = -\sum_k \frac{\partial E}{\partial u_k}\left(\frac{\partial u_k}{\partial x_k}\frac{\partial x_k}{\partial v_m} + \frac{\partial u_k}{\partial y_k}\frac{\partial y_k}{\partial v_m}\right) - \sum_k \frac{\partial E}{\partial v_k}\left(\frac{\partial v_k}{\partial x_k}\frac{\partial x_k}{\partial v_m} + \frac{\partial v_k}{\partial y_k}\frac{\partial y_k}{\partial v_m}\right)$$

$$= \sum_k \delta_{kR}\left(\frac{\partial u_k}{\partial x_k}(-W_{kmI}) + \frac{\partial u_k}{\partial y_k}W_{kmR}\right) + \sum_k \delta_{kI}\left(\frac{\partial v_k}{\partial x_k}(-W_{kmI}) + \frac{\partial v_k}{\partial y_k}W_{kmR}\right) \quad (15)$$

Using the same partial derivatives that helped to establish (7) and (8), and combining (14) and (15), the following expression is obtained similarly for the weight update function (12) using (11):

$$\delta_m = \delta_{mR} + i\delta_{mI} = \sum_k \overline{W}_{km}\,\overline{f}'_m(z_m)\overline{\delta}_k \quad (16)$$

Compared to the fully complex activation representation as $f(z)=u(x,y)+iv(x,y)$, the split complex activation function is a special case and can be represented as $f(z)=u(x)+iv(y)$. This indicates that $u_y=v_x=0$ for the split complex backpropagation algorithm. Removing these zero terms from (10), we obtain the following (complex) weight updates:

$$\Delta W_{nk} = \alpha \overline{X}_{nk}\left(\frac{\partial u_n}{\partial x}\delta_{nR} + i\frac{\partial v_n}{\partial y}\delta_{nI}\right) \quad (17)$$

As before, for output layer neuron, $\delta_n = e_n = d_n - o_n$. For the input and hidden layer,

$$\delta_m = \sum_k \overline{W}_{lm}\left(\frac{\partial u_m}{\partial x_k}\delta_{kR} + i\frac{\partial v_m}{\partial y_k}\delta_{kI}\right). \quad (18)$$

Even though complex multiplication and addition are used throughout the weight update process in equations (17) and (18), the marginal nature of real-to-real and imaginary-to-imaginary interactions that limit the full use of information from the real and imaginary components of the signal is clearly observed from these equations.

## 3. ELEMENTARY TRANSCENDENTAL FUNCTIONS

The following elementary transcendental functions including tanh $z$ are identified to provide adequate nonlinear discriminant as an activation function to restore AM/AM and AM/PM TWTA distortions described in the next section.

| | |
|---|---|
| Circular: | $\tan z,\ \sin z$ |
| Inverse Circular: | $\operatorname{atan} z,\ \operatorname{asin} z,\ \operatorname{acos} z$ |
| Hyperbolic: | $\tanh z,\ \sinh z$ |
| Inverse Hyperbolic: | $\operatorname{atanh} z,\ \operatorname{asinh} z$ |

Figure 1 shows a time-delayed FNN structure that employs these fully complex activation functions. In practice, we note that the existence of singularities that cause these functions to be bounded and analytic *almost everywhere* in **C** hardly poses problems in training and adaptation processes. For example, as shown in Figure 2, tanh $z$ is periodic and also has periodic singularities at every $\pm n\pi i/2$, $n\in\mathbf{N}$, which makes it difficult for large domain applications. However, when the domain of interest is a bounded neighborhood of the unit circle, as is the case in many practical signal processing and communication applications, these singular points hardly pose a problem. From Figure 1, it can be observed that, it is very difficult for the complex weighted sum of time delayed input data to fall exactly on the singular points, and in case it does, a step to handle this case can be included in the implementation.

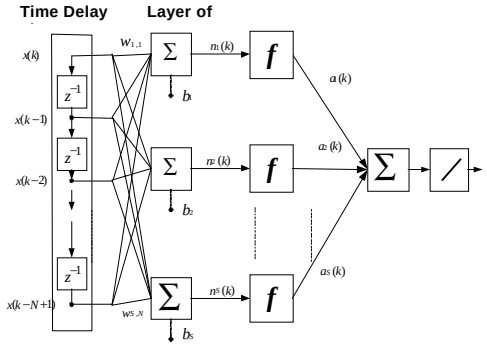


Figure 1. Time-delayed FNN Structure

The characteristics of each transcendental function for suitable applications require further study. For example, as the domain grows, network size and learning rate modification will be needed due to the decreasing dynamic range at the outer region of some functions including asinh $z$ as shown in Figure 3.

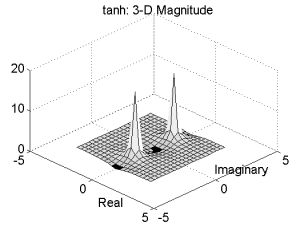


Figure 2. Magnitude of tanh $z$

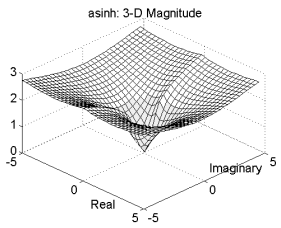


Figure 3. Magnitude of asinh $z$

## 4. NUMERICAL EXAMPLES

The following discrete-input and discrete-output relationship describes the TWTA AM/AM and AM/PM models using Volterra series [9]:

$$x_n = \sum_{k=1}^{\infty}\sum_{n_1}\cdots\sum_{n_{2k-1}} a_{n-n_1}a_{n-n_2}\cdots a_{n-n_k} \bullet a^*_{n-n_{k+1}}\cdots a^*_{n-n_{2k-1}} H^{(2k-1)}_{n_1,\ldots n_{2k-1}} + \nu_n \quad (19)$$

where $\nu_n$ is a complex Gaussian random noise representing the down-link noise, and $a_n$ represents the information symbols, and $H^{(1)}_n, H^{(3)}_{n_1,n_2,n_3}\cdots$ are a set of complex Volterra series coefficients that describe the effect of the nonlinear channel on symbols $a_n$. The reduced set of Volterra coefficients is obtained by assuming constant modulus modulation as in [9]. But, they are used here for the equalization of non-constant modulus quadrature amplitude modulated (QAM) signals transmitted through TWTA AM/AM and AM/PM channels to demonstrate the capability of fully complex backpropagation using the elementary transcendental

functions described in the previous section. Figure 4 shows a 16-QAM constellation with 21 dB signal-to-noise ratio (Eb/No) that reaches the saturation level of unit amplitude at the outer most corner points. Figure 5 shows the AM/AM and AM/PM impact on 16-QAM symbols transmitted through TWTA.

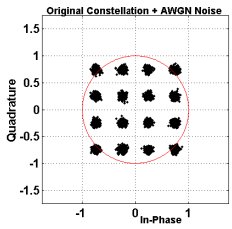


Figure 4. Input 16-QAM constellation

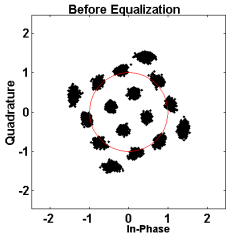


Figure 5. Received 16-QAM constellation after TWTA

Figures 6 and 7 show Eb/No versus symbol error rate (SER) performances of thirteen schemes discussed in this paper. They include the fully complex activation functions of Georgiou and Koutsougeras. [2] and Hirose [3], along with the nine transcendental functions we introduced in Section 3, the split complex hyperbolic tangent, and the CLMS. Figure 6 shows relatively poor performers that could not achieve SER below 0.001 at 14 dB Eb/No while Figure 7 includes better performers. Former group includes the activation functions of [3] and [2], tanh *z*, sin *z*, sinh *z*, atanh *z*, and acos *z*, while the latter group includes the split-tanh *x*, tan *z*, atan *z*, asin *z*, CLMS, and asinh *z*, respectively, in the order of worse to better SER performance.

Note that the split-tanh *x* scheme did not perform well compared to the four fully complex activation functions and also with respect to the CLMS. Both figures show that all equalizers failed to approach the performance of the pure additive white Gaussian noise (AWGN) channel. This is because of the difficulty of estimating independent Gaussian amplitude and uniform phase noise distributions that tends to spread out the constellation after the equalization. It is also worth noting that CLMS performed well under this non-saturated and almost linear channel distortion environment as it uses the information from the real and complex components effectively.

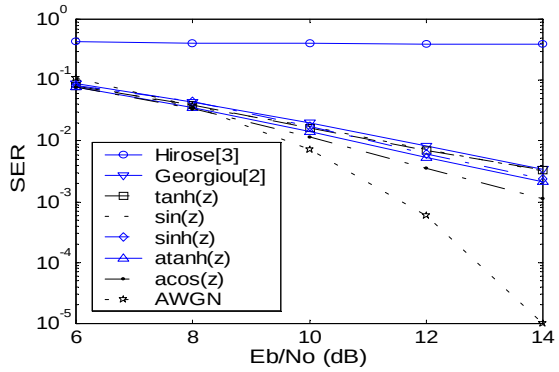


Figure 6. Eb/No vs. SER of relatively poor performers

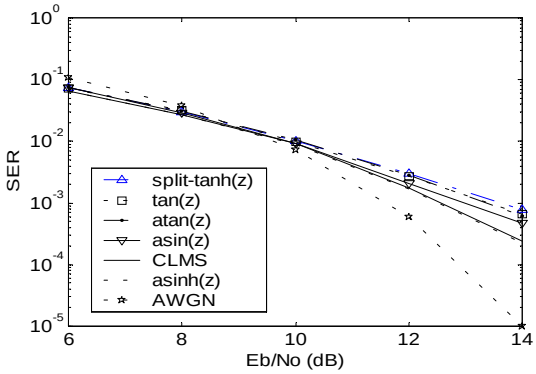


Figure 7. Eb/No vs. SER of better performers

The advantage of asinh z activation function over CLMS is demonstrated in Figure 8 under a highly nonlinear saturated TWTA channel environment shown in Figure 9. Note that the SER curves for this case cannot significantly improve the performance at high Eb/No's as the nonlinear distortion is very high.

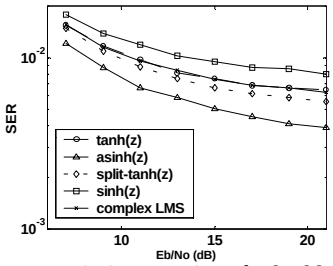


Figure 8. SNR vs. SER for highly nonlinear saturated TWTA

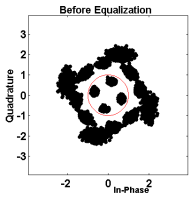


Figure 9.Received saturated 16-QAM constellation

## 5. REFERENCES


[1] Silverman H., *Complex Variables*, Houghton, Newark, USA, 1975.

[2] Georgiou G. and Koutsougeras C., "Complex Backpropagation", *IEEE Trans. on Circuits and Systems II*, Vol. 39, No.5pp. 330 – 334, 1992.

[3] Hirose A., "Continuous Complex-Valued Back-propagation Learning", *Electronics Letters*, Vol.28, No. 20, pp. 1854 – 1855, September 24, 1992.

[4] Leung H. and Haykin S., "The Complex Backpropagation Algorithm", *IEEE Trans. on Signal Proc.*, Vol. 3, No. 9, pp. 2101 – 2104, 1991.

[5] Benvenuto N., Marchesi M., Piazza F., and Uncini A., "Non Linear Satellite Radio Links Equalized using Blind Neural Networks", *Proc. of ICASSP* Vol. 3, pp. 1521 –1524, 1991.

[6] Benvenuto N. and Piazza F., "On the Complex Backpropagation Algorithm", *IEEE Trans. on Signal Processing*, Vol. 40, No. 4, pp. 967 – 969, 1992.

[7] You C. and Hong D., "Nonlinear Blind Equalization Schemes using Complex-Valued Multilayer Feedforward Neural Networks", *IEEE Trans. on Neural Networks*, Vol. 9, No. 6, pp. 1442 – 1455, 1998.

[8] Kim T. and Adali T., "Fully Complex Backpropagation for Constant Envelop Signal Processing", *Proc. of IEEE Workshop on Neural Networks for Sig. Proc.*, pp. 231-240, Sydney, Dec. 2000.

[9] Benedetto S. and Biglieri E. "Nonlinear Equalization of Digital Satellite Channels", *IEEE Jour. on SAC..*, vol. SAC-1., pp. 57-62, 1983.