Journal of VLSI Signal Processing 32, 29–43, 2002

Fully Complex Multi-Layer Perceptron Network for Nonlinear Signal Processing

TAEHWAN KIM
Center for Advanced Aviation System Development, The MITRE Corporation, M/S N670, 7515 Colshire Drive, McLean, Virginia 22102, USA; Information Technology Laboratory, Department of Computer Science and Electrical Engineering, University of Maryland Baltimore County, Baltimore, Maryland 21250, USA

TÜLAY ADALI
Information Technology Laboratory, Department of Computer Science and Electrical Engineering, University of Maryland Baltimore County, Baltimore, Maryland 21250, USA

Received July 2, 2001; Revised November 14, 2001; Accepted November 19, 2001

Abstract. Designing a neural network (NN) to process complex-valued signals is a challenging task since a complex nonlinear activation function (AF) cannot be both analytic and bounded everywhere in the complex plane $\mathbb{C}$. To avoid this difficulty, '*splitting*', i.e., using a pair of real sigmoidal functions for the real and imaginary components has been the traditional approach. However, this 'ad hoc' compromise to avoid the unbounded nature of nonlinear complex functions results in a nowhere analytic AF that performs the error back-propagation (BP) using the split derivatives of the real and imaginary components instead of relying on well-defined fully complex derivatives. In this paper, a fully complex multi-layer perceptron (MLP) structure that yields a simplified complex-valued back-propagation (BP) algorithm is presented. The simplified BP verifies that the fully complex BP weight update formula is the complex conjugate form of real BP formula and the split complex BP is a special case of the fully complex BP. This generalization is possible by employing elementary transcendental functions (ETFs) that are *almost everywhere* (a.e.) bounded and analytic in $\mathbb{C}$. The properties of fully complex MLP are investigated and the advantage of ETFs over *split* complex AF is shown in numerical examples where nonlinear magnitude and phase distortions of non-constant modulus modulated signals are successfully restored.

Keywords: nonlinear adaptive signal processing, fully complex neural network, split complex neural network, elementary transcendental functions, bounded almost everywhere, analytic almost everywhere

1. Introduction

The main reason for the difficulty in finding a nonlinear complex activation function in MLP design is the conflict between the boundedness and the differentiability of complex functions in the whole complex plane $\mathbb{C}$. As stated by Liouvilles theorem, *a bounded entire function must be a constant in* $\mathbb{C}$, where an entire function is defined as analytic, i.e., differentiable at every point $z \in \mathbb{C}$ [1]. Therefore, we cannot find an analytic complex nonlinear activation function that is bounded everywhere on the entire $\mathbb{C}$.

In general, the fact that many real-valued mathematical objects are best understood when they are considered as embedded in the complex plane (e.g., the Fourier transform) provides the motivation to develop a neural network (NN) structure that uses well-defined fully complex nonlinear activation functions [2]. However, due to the Liouville's theorem, the complex NN has been subjected to the common view that it

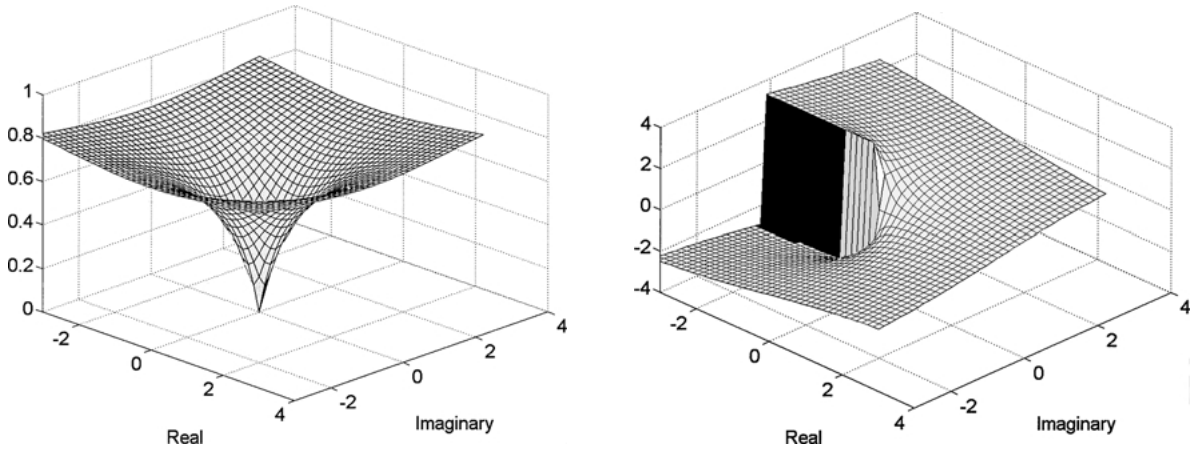
has to live with a compromised employment of the *non-analytic but bounded* nonlinear AF for the stability of the system [3, 4]. We employ a subset of ETFs derivable from the function $f(z) = e^z$ that is an entire function since $f(z) = f'(z) = e^z$. These ETFs are entire but bounded *almost everywhere*, i.e., are only unbounded on a set of points having zero measure. Since these functions are entire, they are also conformal and provide well-defined derivatives while still allowing the complex MLP to converge with probability 1, which is sufficient in practical signal processing applications. Also, among these ETFs, circular and hyperbolic ETFs such as $\tan z$ and $\tanh z$ are bilinear (i.e., Möbius) transforms and lend themselves to fixed-point analysis as discussed in [5]. Note that, however, inverse ETFs such as $\operatorname{arctanh} z$ are not bilinear transforms, as we discuss in Section 4.

To overcome the unbounded nature of analytic functions in $\mathbb{C}$, earlier fully complex activation functions, proposed by Georgiou and Koutsougeras [3] (given in Eq. (1)) and Hirose [6] (given in Eq. (2)) have normalized or scaled the amplitude of complex signals:

$$f(z) = z/(c + |z|/r) \tag{1}$$

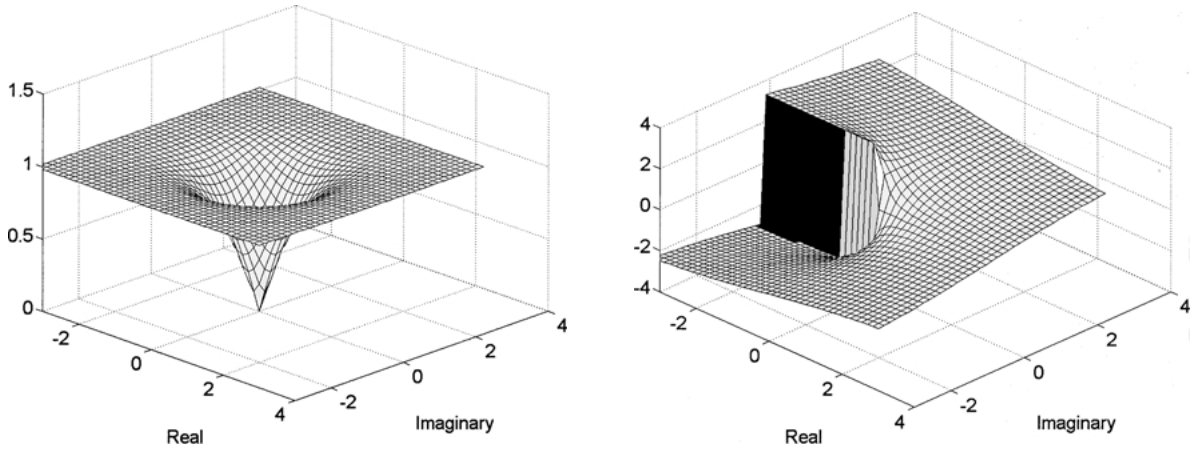
$$f(s\exp[i\beta]) = \tanh(s/m)\exp[i\beta] \tag{2}$$

where, c, r, and m are real positive constants and $z = s\exp[i\beta]$. Figure 1(a) and (b) show the magnitude and phase response of Eq. (1) while Fig. 2(a) and (b) show those of Eq. (2).



(a) (b)

Figure 1. (a) Magnitude of $f(z) = z/(c + |z|/r)$. (b) Phase of $f(z) = z/(c + |z|/r)$.



(a) (b)

Figure 2. (a) Magnitude of $f(s\exp[i\beta]) = \tanh(s/m)\exp[i\beta]$. (b) Phase of $f(s\exp[i\beta]) = \tanh(s/m)\exp[i\beta]$.

However, these functions preserve the phase, thus are less efficient in learning the nonlinear phase variations between the input and the targets than those complex activation functions with nonlinear phase mapping capability. A subset of ETFs introduced in Section 4 possesses amplitude and phase nonlinear mapping capability and are bounded and analytic a.e. in a bounded domain of $\mathbb{C}$. The numerical examples included in Section 5 show that these ETFs perform better in restoring nonlinear amplitude and phase distortion in a noisy environment. More specifically, the MLP structure using the AF of Eq. (1) performs poorly in restoring nonlinear distortion of nonconstant modulus signals due to its inflexible radial mapping characteristics as shown in the numerical examples included in Section 5. Hirose's function given in Eq. (2) performs relatively better than Eq. (1), but still worse than most of the ETFs we employ.

More traditional approach for processing complex signals with MLPs is to use a 'split' complex AF where a pair of conventional real-valued AFs marginally process the in-phase (I) and quadrature (Q) components [4, 7–13]. While this approach can avoid the unboundedness of fully complex AFs in view of Liouville's theorem, the split complex AF cannot be analytic. The split approach typically employs a pair of real-valued functions $f_R(x) = f_I(x) = \tanh x$, $x \in \mathbb{R}$, as a nonlinear complex AF in the hidden layer [7–13], as shown in Eq. (3).

$$f(z) = f_R(\mathrm{Re}(z)) + if_I(\mathrm{Im}(z)) \tag{3}$$

where $z = X^T W$, $X \in \mathbb{C}^n$ is the n-dimensional complex input, and $W \in \mathbb{C}^n$ is the weight vector. This approach has been shown to be adequate in providing equalization of nonlinear satellite communication channels. But as is shown in the numerical examples in Section 5, fully complex ETFs we introduce outperform the split complex AF in bit error rate performance in our equalization experiments. The well-defined first order derivatives of complex ETFs also satisfy the *Cauchy-Riemann* equations, therefore allows for simplification of the fully complex BP that is given in [3]. The simplification verifies that the weight update formula of fully complex BP is a complex conjugate form of the real weight update formula. Also, while the complex ETFs provide well-defined derivatives for optimization of the fully complex BP algorithm, the derivatives in split complex BP cannot fully exploit the correlation between the real and imaginary components of weighted sum of input vectors. Figure 3(a) and (b) show the magnitude and phase of split tanh x function, respectively.

In Section 2 of the paper, the desired properties of fully complex AFs stated as necessary in [3] and [4] are reduced and relaxed into a single condition. In Section 3, a more compact version of the fully complex BP algorithm than the one derived in [3] is given followed by the comparison between fully and split complex BP algorithms. It is shown that the split complex BP is a special case of the fully complex BP algorithm that we derive. In Section 4, nine ETFs that satisfy the modified desired properties of a fully complex activation function are presented with

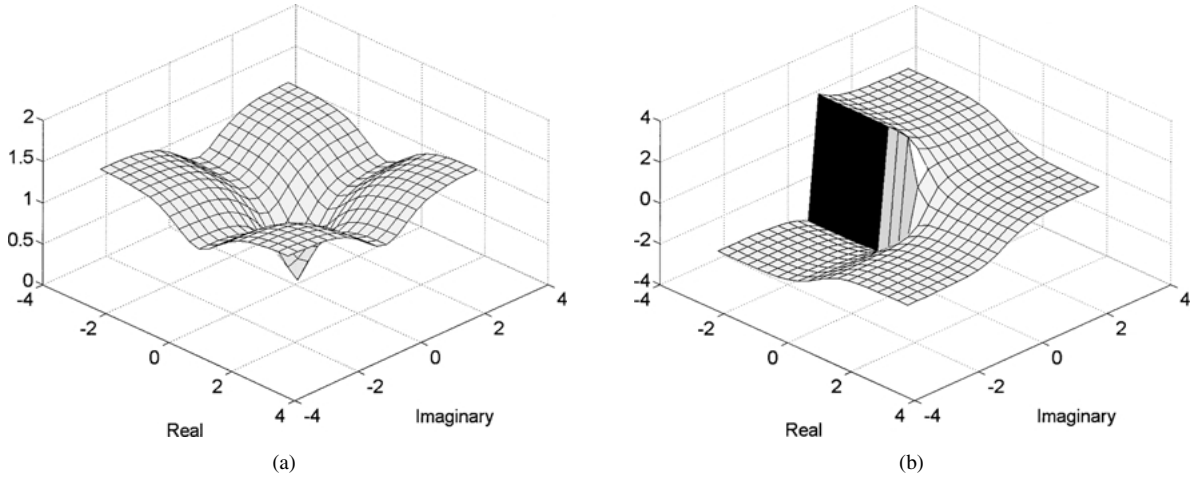


Figure 3. (a) Magnitude of split tanh x. (b) Phase of split tanh x.

a and discussion of their characteristics. The performances of the BP algorithm using nine ETFs along with earlier AFs shown in Eqs. (1) and (2) are compared with the split complex and complex least mean square (CLMS) equalizers in numerical examples in Section 5. Here, the *Volterra* series nonlinear satellite Traveling Wave Tube Amplifier (TWTA) model [12] exhibiting amplitude-to-amplitude (AM/AM) and amplitude-to-phase (AM/PM) distortion is used as the nonlinear channel model. Non-constant modulus signals are distorted in mild and severe nonlinear conditions based on two levels of amplifier saturation. For better evaluation of different MLPs, multiple levels of training stopping criteria are used to compare the closest-to-the-best performances of each training scheme and the learning behavior.

2. Desired Properties of a Fully Complex Activation Function

The unbounded nature of nonlinear analytic functions in $\mathbb{C}$ triggered the definition of "desirable" properties for the complex AFs. The first attempt was made by Georgiou and Koutsougeras [3] who identified five desirable properties of a fully complex AF $f(z)$:

1. $f(z) = u(x, y) + iv(x, y)$ is nonlinear in x and y
2. $f(z)$ is bounded
3. The partial derivatives u_x, u_y, v_x, and v_y exist and are bounded
4. $f(z)$ is not entire
5. $u_x v_y \neq v_x u_y$.

Note that by Liouville's theorem, the second and the fourth conditions are redundant, i.e., a bounded nonlinear function in $\mathbb{C}$ cannot be entire. Using this fact, You and Hong [4] reduced the above conditions to four conditions given below and introduced a split complex AF for the Quadrature Amplitude Modulation (QAM) signal equalization problem they considered:

1. $f(z) = u(x, y) + iv(x, y)$ is nonlinear in x and y
2. For the stability of a system, $f(z)$ should have no singularities and be bounded for all z in a bounded set
3. The partial derivatives u_x, u_y, v_x, and v_y should be continuous and bounded
4. $u_x v_y \neq v_x u_y$. If not, then $f(z)$ is not a suitable activation function except in the following cases:
   (i) $u_x = v_x = 0$, and $u_y \neq 0$, $v_y \neq 0$,
   (ii) $u_y = v_y = 0$, and $u_x \neq 0$, $v_x \neq 0$.

Note that both sets of conditions above are emphasizing boundedness of an AF and its partial derivatives, even when the function is defined in a local domain of interest. The ETFs that we propose violate the second and third boundedness requirements of both sets of conditions above. But these violations only occur on a set of points with zero-measures, i.e., they are bounded and analytic. a.e. This property results in convergence with probability 1 that has no practical significance in engineering applications. More importantly, these functions outperform other linear complex schemes and AFs given in Eqs. (1) and (2) in learning nonlinear distortion once the domain is bounded as boundedness naturally sets the limit on the magnitude of range (see the discussions on the characteristics of these elementary transcendental functions in Section 4).

The last restriction $u_x v_y \neq v_x u_y$ was originally imposed to guarantee continuous learning, i.e., to prevent a situation where a non-zero activation function input forces the gradient of the error function (with respect to the complex weight) to become zero. We also note that this condition is unnecessary for fully complex ETFs because they satisfy the *Cauchy-Riemann* equations shown in Eq. (4).

For non-continuous functions, *Cauchy-Riemann* equations are the necessary condition for a complex function to be analytic at a point $z \in \mathbb{C}$. They can be written by noting that the partial derivatives of $f(z) = u(x, y) + iv(x, y)$, where $z = x + iy$, along the real and imaginary axes should be equal:

$$f'(z) = u_x + iv_x = v_y - iu_y. \tag{4}$$

Equating the real and imaginary parts above, we obtain the *Cauchy-Riemann* equations: $u_x = v_y$, $v_x = -u_y$. Once *Cauchy-Riemann* equations are met, the last restriction $u_x v_y \neq v_x u_y$ can be replaced by $u_x v_y \neq v_x u_y \Rightarrow u_x^2 \neq -u_y^2$, which is always true except when $u_x = u_y = 0$. This result also leads to $v_x = v_y = 0$, since all four partial derivatives are real. This only happens when a local extremum occurs where the learning process should naturally stop. Therefore, the last condition is only necessary for non-analytic split complex functions but not for analytic fully complex functions.

Consequently, since there is no interest in using a linear function as an activation function and the partial derivatives are well defined to satisfy the *Cauchy-Riemann* equations, the desirable properties of a fully complex activation function is reduced and relaxed as follows:

In a bounded domain of complex plane $\mathbb{C}$, a fully complex nonlinear activation function $f(z)$ needs to be analytic and bounded a.e.

3. Complex Back-Propagation

Assuming that the fully complex AF satisfies the condition above such that it is analytic a.e., the *Cauchy-Riemann* equations can be used to simplify the fully complex BP algorithm derived in [3] as shown in this section. The simplification uses the following concise expression obtained from Eq. (4):

$$f'(z) = f_x = -if_y \tag{5}$$

where f_x and f_y are partial derivates with respect to x and y respectively.

If the fully complex activation function is $f(z) = u(x, y) + iv(x, y)$, then the sum-squared error at the output layer can be written as

$$E = \frac{1}{2}\sum_n |e_n|^2, \quad \text{where } e_n = d_n - o_n \tag{6}$$

$$o_n \equiv f(z_n) \equiv u_n + iv_n, z_n \equiv x_n + iy_n \equiv \sum_k W_{nk}X_{nk}$$
$$= \sum_k \left(W_{nk}^R + iW_{nk}^I\right)\left(X_{nk}^R + iX_{nk}^I\right) \tag{7}$$

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 $\partial E/\partial W_{nk}$. Since the cost function is a real function of a complex variable, the gradient of the error function with respect to the real and imaginary components of W_{nk} can be written as

$$\frac{\partial E}{\partial W_{nk}} \equiv \nabla_{W_{nk}} E = \frac{\partial E}{\partial W_{nk}^R} + i\frac{\partial E}{\partial W_{nk}^I}, \tag{8}$$

and using the chain rule,

$$\frac{\partial E}{\partial W_{nk}^R} = \frac{\partial E}{\partial u_n}\left(\frac{\partial u_n}{\partial x_n}\frac{\partial x_n}{\partial W_{nk}^R} + \frac{\partial u_n}{\partial y_n}\frac{\partial y_n}{\partial W_{nk}^R}\right) + \frac{\partial E}{\partial v_n}\left(\frac{\partial v_n}{\partial x_n}\frac{\partial x_n}{\partial W_{nk}^R} + \frac{\partial v_n}{\partial y_n}\frac{\partial y_n}{\partial W_{nk}^R}\right) \tag{9}$$

$$\frac{\partial E}{\partial W_{nk}^I} = \frac{\partial E}{\partial u_n}\left(\frac{\partial u_n}{\partial x_n}\frac{\partial x_n}{\partial W_{nk}^I} + \frac{\partial u_n}{\partial y_n}\frac{\partial y_n}{\partial W_{nk}^I}\right) + \frac{\partial E}{\partial v_n}\left(\frac{\partial v_n}{\partial x_n}\frac{\partial x_n}{\partial W_{nk}^I} + \frac{\partial v_n}{\partial y_n}\frac{\partial y_n}{\partial W_{nk}^I}\right). \tag{10}$$

Defining $\delta_n^R \equiv -\partial E/\partial u_n$ and $\delta_n^I \equiv -\partial E/\partial v_n$, and using the following partial derivatives identifiable from Eq. (7),

$$\frac{\partial x_n}{\partial W_{nk}^R} = X_{nk}^R; \quad \frac{\partial y_n}{\partial W_{nk}^R} = X_{nk}^I;$$
$$\frac{\partial x_n}{\partial W_{nk}^I} = -X_{nk}^I; \quad \frac{\partial y_n}{\partial W_{nk}^I} = X_{nk}^R$$

Eqs. (9) and (10) can be simplified as

$$\frac{\partial E}{\partial W_{nk}^R} = -\delta_n^R\left(\frac{\partial u_n}{\partial x_n}X_{nk}^R + \frac{\partial u_n}{\partial y_n}X_{nk}^I\right) - \delta_n^I\left(\frac{\partial v_n}{\partial x_n}X_{nk}^R + \frac{\partial v_n}{\partial y_n}X_{nk}^I\right) \tag{11}$$

$$\frac{\partial E}{\partial W_{nk}^I} = -\delta_n^R\left(\frac{\partial u_n}{\partial x_n}\left(-X_{nk}^I\right) + \frac{\partial u_n}{\partial y_n}X_{nk}^R\right) - \delta_n^I\left(\frac{\partial v_n}{\partial x_n}\left(-X_{nk}^I\right) + \frac{\partial v_n}{\partial y_n}X_{nk}^R\right). \tag{12}$$

Combining Eqs. (11) and (12), the gradient of the error function can be simplified as

$$\nabla_{W_{nk}} E = -\bar{X}_{nk}\left\{\left(\frac{\partial u_n}{\partial x_n} + i\frac{\partial u_n}{\partial y_n}\right)\delta_n^R + \left(\frac{\partial v_n}{\partial x_n} + i\frac{\partial v_n}{\partial y_n}\right)\delta_n^I\right\}. \tag{13}$$

Note that Eq. (13) was first established in [3], but further applying the *Cauchy-Riemann* equations as shown in Eq. (14), the compact form of fully complex BP algorithm is obtained in Eq. (15).

$$\begin{cases} \dfrac{\partial u_n}{\partial x_n} + i\dfrac{\partial u_n}{\partial y_n} = \dfrac{\partial u_n}{\partial x_n} - i\dfrac{\partial v_n}{\partial x_n} = \dfrac{\partial \bar{f}_n}{\partial x_n} = \bar{f}'(z_n) \\ \dfrac{\partial v_n}{\partial x_n} + i\dfrac{\partial v_n}{\partial y_n} = -\dfrac{\partial u_n}{\partial y_n} + i\dfrac{\partial v_n}{\partial y_n} \\ \quad = -\left(\dfrac{\partial u_n}{\partial y_n} - i\dfrac{\partial v_n}{\partial y_n}\right) = -\dfrac{\partial \bar{f}_n}{\partial y_n} = i\bar{f}'(z_n) \end{cases} \tag{14}$$

$$\nabla_{W_{nk}} E = -\bar{X}_{nk}\left(\bar{f}'(z_n)\delta_n^R + i\bar{f}'(z_n)\delta_n^I\right) = -\bar{X}_{nk}\bar{f}'(z_n)\delta_n \tag{15}$$

It is worth noting that except the conjugate on X_{nk} and $f'(z_n)$ terms, Eq. (15) is identical to the gradient of

the error function in the real-valued backpropagation algorithm.

Complex weight update ΔW_{nk} is proportional to the negative gradient:

$$\Delta W_{nk} = \alpha \bar{X}_{nk} \bar{f}'(z_n)\delta_n \tag{16}$$

where, α can be either a real, imaginary, or complex learning rate. The performance differences with each type of learning rate and the choice for a given problem is an interesting problem that needs to be investigated. When the complex weight belongs to an output neuron: $\delta_n = e_n = d_n - o_n$ and when W_{mk} belongs to the m-th hidden layer, the net input z_m to neuron m is the same as Eq. (7):

$$z_m = x_m + iy_m = \sum_k (u_k + iv_k)\left(W_{mk}^R + iW_{mk}^I\right) \tag{17}$$

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

$$\begin{aligned}
\delta_m^R &= -\frac{\partial E}{\partial u_m} \\
&= -\sum_k \frac{\partial E}{\partial u_k}\left(\frac{\partial u_k}{\partial x_k}\frac{\partial x_k}{\partial u_m} + \frac{\partial u_k}{\partial y_k}\frac{\partial y_k}{\partial u_m}\right) \\
&\quad -\sum_k \frac{\partial E}{\partial v_k}\left(\frac{\partial v_k}{\partial x_k}\frac{\partial x_k}{\partial u_m} + \frac{\partial v_k}{\partial y_k}\frac{\partial y_k}{\partial u_m}\right) \\
&= \sum_k \delta_k^R\left(\frac{\partial u_k}{\partial x_k}W_{mk}^R + \frac{\partial u_k}{\partial y_k}W_{mk}^I\right) \\
&\quad + \sum_k \delta_k^I\left(\frac{\partial v_k}{\partial x_k}W_{mk}^R + \frac{\partial v_k}{\partial y_k}W_{mk}^I\right).
\end{aligned} \tag{18}$$

Similarly,

$$\begin{aligned}
\delta_m^I &= -\frac{\partial E}{\partial v_m} \\
&= -\sum_k \frac{\partial E}{\partial u_k}\left(\frac{\partial u_k}{\partial x_k}\frac{\partial x_k}{\partial v_m} + \frac{\partial u_k}{\partial y_k}\frac{\partial y_k}{\partial v_m}\right) \\
&\quad -\sum_k \frac{\partial E}{\partial v_k}\left(\frac{\partial v_k}{\partial x_k}\frac{\partial x_k}{\partial v_m} + \frac{\partial v_k}{\partial y_k}\frac{\partial y_k}{\partial v_m}\right) \\
&= \sum_k \delta_k^R\left(\frac{\partial u_k}{\partial x_k}\left(-W_{mk}^I\right) + \frac{\partial u_k}{\partial y_k}W_{mk}^R\right) \\
&\quad + \sum_k \delta_k^I\left(\frac{\partial v_k}{\partial x_k}\left(-W_{mk}^I\right) + \frac{\partial v_k}{\partial y_k}W_{mk}^R\right).
\end{aligned} \tag{19}$$

Using the same partial derivatives that helped to establish Eqs. (11) and (12), and combining Eqs. (18) and (19), the following expression is obtained similarly for the weight update function:

$$\delta_m = \delta_m^R + i\delta_m^I = \sum_k \bar{W}_{mk}\bar{f}'(z_m)\,\delta_k \tag{20}$$

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 Eq. (13), we obtain the following (complex) weight updates:

$$\nabla W_{nk} = \alpha \bar{X}_{nk}\left(\frac{\partial u_n}{\partial x_n}\delta_n^R + i\frac{\partial v_n}{\partial y_n}\delta_n^I\right). \tag{21}$$

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

$$\delta_m = \sum_k \bar{W}_{mk}\left(\frac{\partial u_k}{\partial x_m}\delta_k^R + i\frac{\partial v_k}{\partial y_m}\delta_k^I\right). \tag{22}$$

Even though complex multiplication and addition are used throughout the weight update process in Eqs. (21) and (22), 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 in these equations. Note that the split complex weight update form shown in Eq. (22) can be also simplified to Eq. (20) which was first described by Leung and Haykin in [7].

4. Elementary Transcendental Functions

The following elementary transcendental functions including $\tanh z$ are identified to provide adequate nonlinear discriminant as an activation function and they satisfy the generalized condition we give in Section 2.

- Circular functions

$$\tan z = \frac{e^{iz} - e^{-iz}}{i(e^{iz} + e^{-iz})}, \quad \left\{\frac{d}{dz}\tan z = \sec^2 z\right\},$$
$$\sin z = \frac{e^{iz} - e^{-iz}}{2i}, \quad \left\{\frac{d}{dz}\sin z = \cos z\right\}$$

- Inverse circular functions

$$\arctan z = \int_0^z \frac{dt}{1+t^2}, \quad \left\{\frac{d}{dz}\arctan z = \frac{1}{1+z^2}\right\},$$
$$\arcsin z = \int_0^z \frac{dt}{(1-t)^{1/2}},$$
$$\left\{\frac{d}{dz}\arcsin z = (1-z^2)^{-1/2}\right\},$$
$$\arccos z = \int_z^1 \frac{dt}{(1-t^2)^{1/2}},$$
$$\left\{\frac{d}{dz}\arccos z = -(1-z^2)^{-1/2}\right\}$$

- Hyperbolic functions

$$\tanh z = \frac{\sinh z}{\cosh z} = \frac{e^z - e^{-z}}{e^z + e^{-z}}, \quad \left\{\frac{d}{dz}\tanh z = \sec h^2 z\right\},$$
$$\sinh z = \frac{e^z - e^{-z}}{2}, \quad \left\{\frac{d}{dz}\sinh z = \cosh z\right\}$$

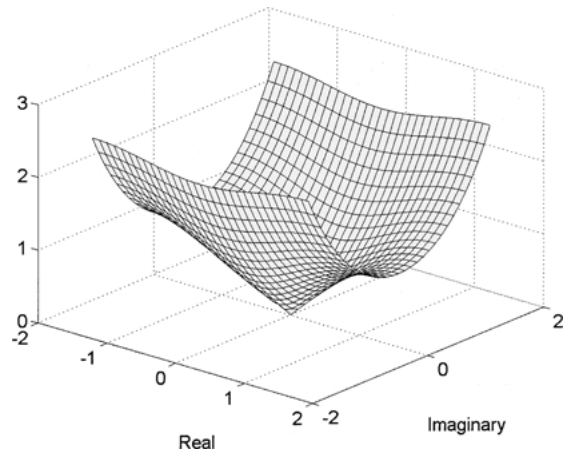
- Inverse hyperbolic functions

$$\operatorname{arctanh} z = \int_0^z \frac{dt}{1-t^2}, \quad \left\{\frac{d}{dz}\operatorname{arctanh} z = (1-z^2)^{-1}\right\},$$
$$\operatorname{arcsinh} z = \int_0^z \frac{dt}{(1+t^2)^{1/2}},$$
$$\left\{\frac{d}{dz}\operatorname{arcsinh} z = (1+z^2)^{-1}\right\}$$

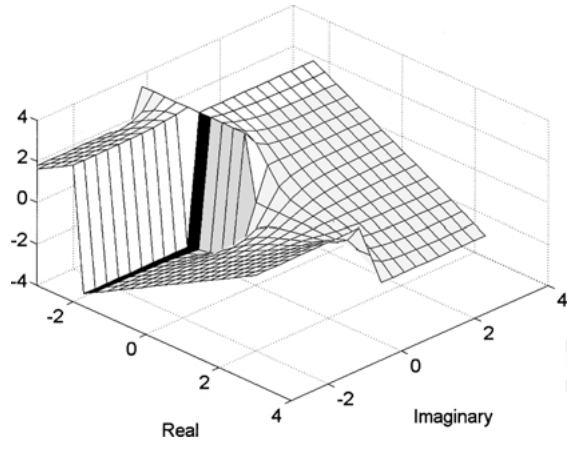
We employ these activation functions in an MLP structure that uses a tap delayed input line proportional to the expected correlation span in the observations (channel outputs). Figures 4 through 12 show the magnitude and phase of these elementary transcendental functions.

Note the periodicity and singularity of $\tanh z$ at every $(1/2+n)\pi i, n \in \mathbb{N}$, in Fig. 6, while Figs. 5, 7, and 11 show singularities of $\tan z$, $\arctan z$, and $\operatorname{arctanh} z$ at $(1/2+n)\pi, \pm i$ and ± 1, respectively. Usually, these singular points and discontinuities at non-zero points do not pose a problem in training when the domain of interest is bounded within a circle of radius $\pi/2$. If the domain is larger including these irregular points, the training process tends to become more sensitive to the size of the learning rate and poses a penalty when a fixed learning rate is used. Also, it is observed that the initial random weights need to be bounded in a small radius, typically below 0.1 to avoid oscillation in the gradient based updates.

Unlike $\tan z$ and $\tanh z$ functions with pointwise singularity, Figs. 8 and 12 show that $\arcsin z$ and $\operatorname{arcsinh} z$ functions are not continuous and therefore not analytic along a portion of the real or the imaginary axis. However, it is most radially symmetric in magnitude, while Fig. 9 shows that $\arccos z$ function is asymmetric along real axis and has a discontinuous zero point at $z = 1$. The split-tanh x function in Fig. 3 shows that it is not quite radially symmetric because of the valleys along the real and imaginary axes. Note that, these functions have decreasing rates of amplitude growth as we move away from the origin. Because of the decreasing rate of amplitude growth, the application of scaling coefficient in the activation function may become necessary when the data shows large dynamic range. Also

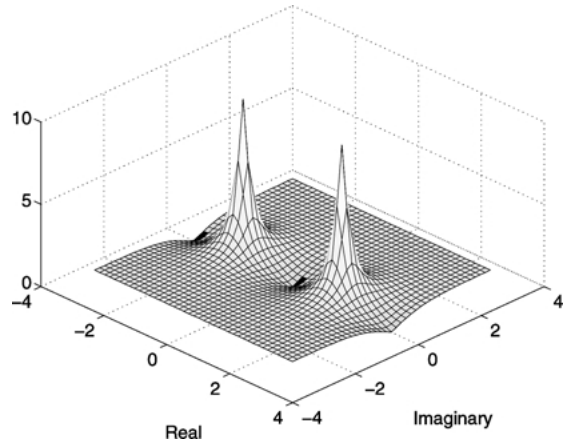


(a)

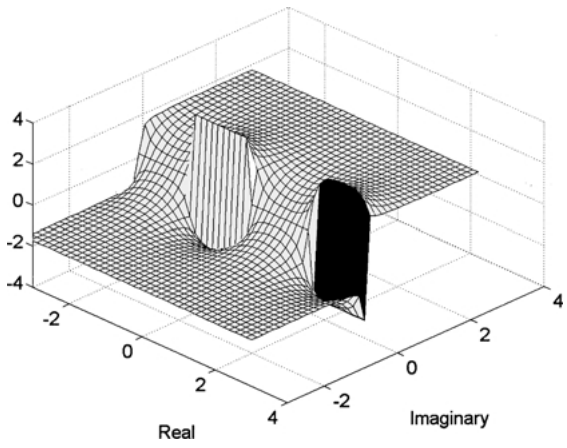


(b)

Figure 4. (a) Magnitude of $\sin z$. (b) Phase of $\sin z$.

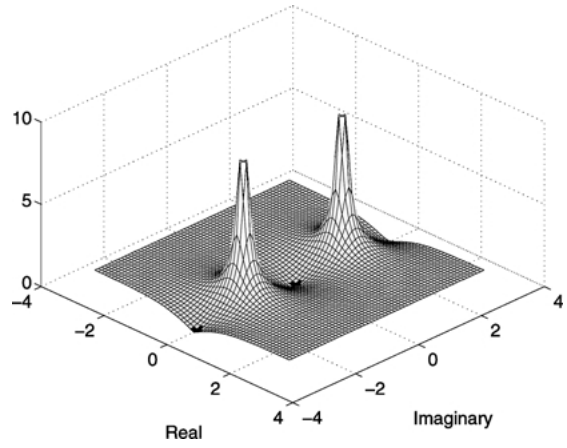


(a)

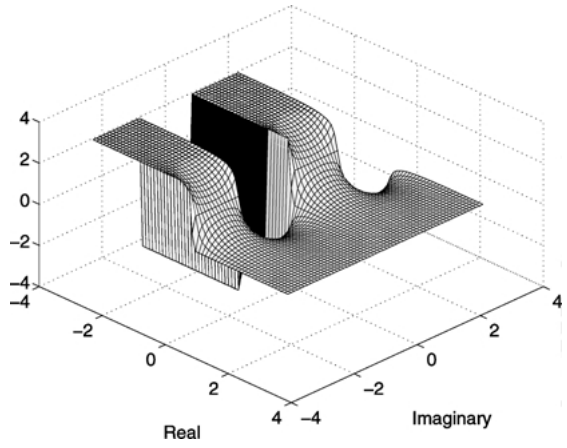


(b)

Figure 5. (a) Magnitude of tan z. (b) Magnitude of tan z.

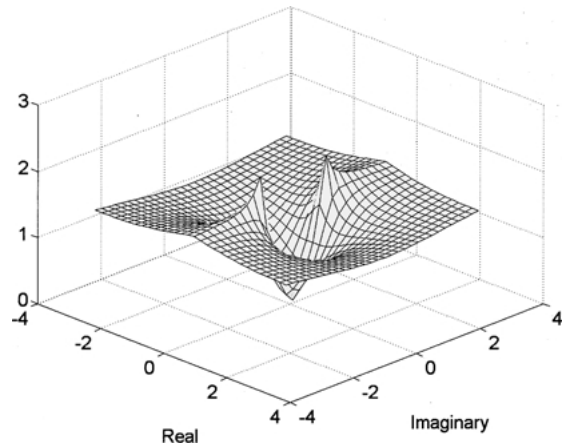


(a)

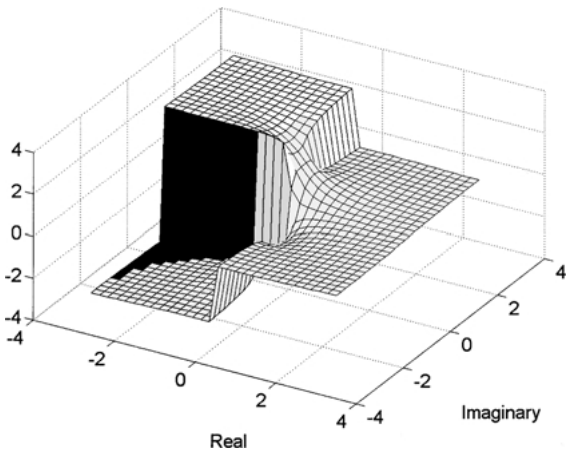


(b)

Figure 6. (a) Magnitude of tanh z. (b) Phase of tanh z.

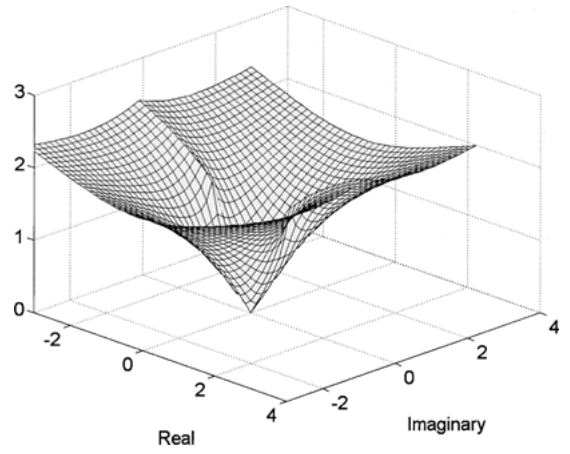


(a)



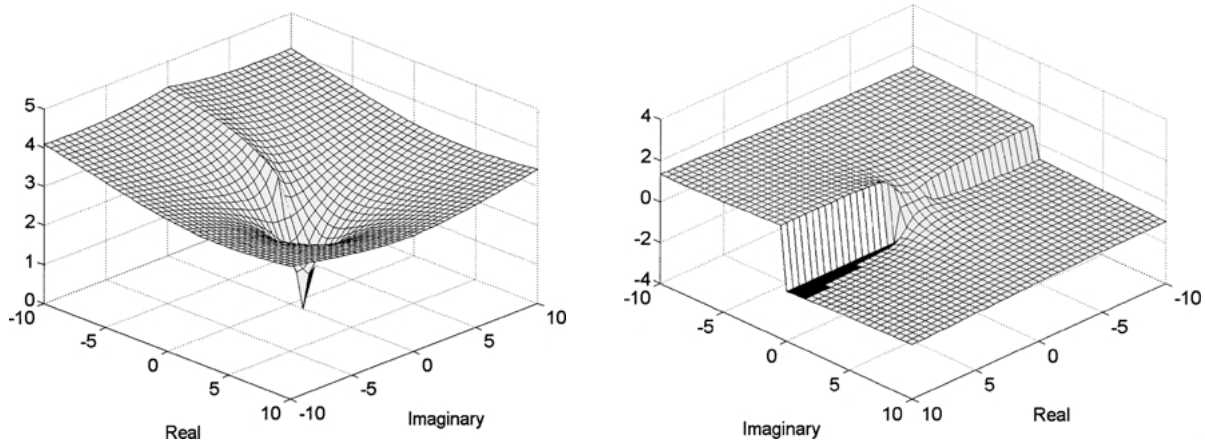
(b)

Figure 7. (a) Magnitude of arctan z. (b) Phase of arctan z.



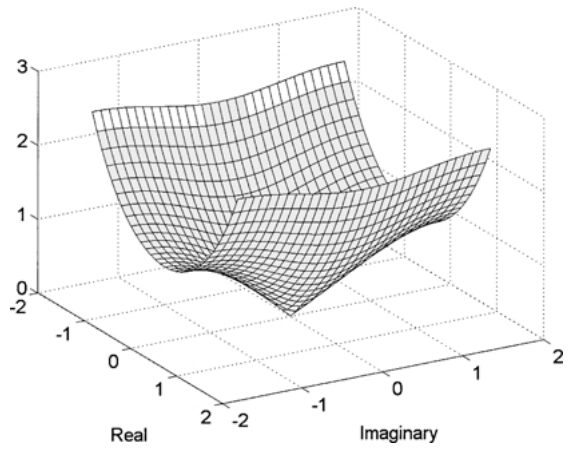
(a) (b)

Figure 8. (a) Magnitude of arcsin z. (b) Phase of arcsin z.



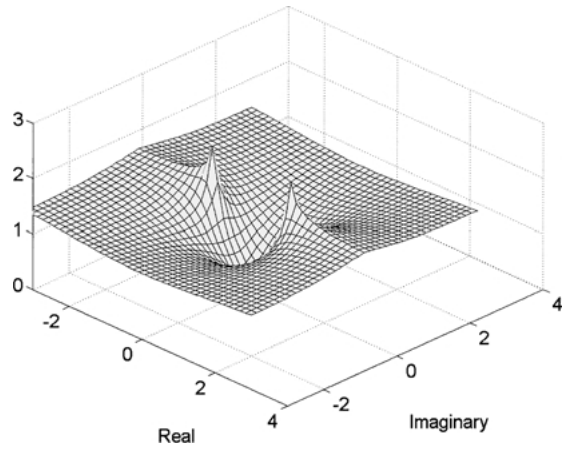
(a) (b)

Figure 9. (a) Magnitude of arccos z. (b) Phase of arccos z.

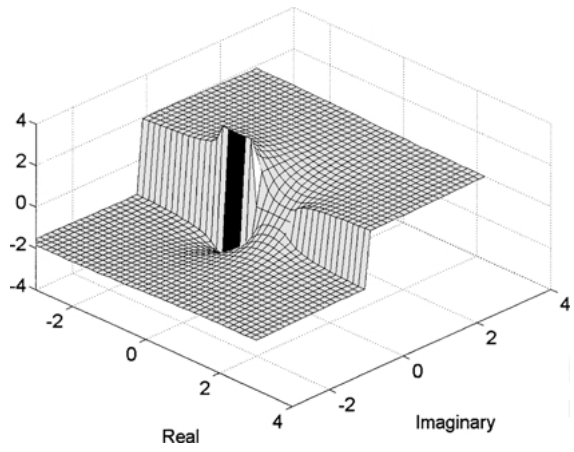


(a) (b)

Figure 10. (a) Magnitude of sinh z. (b) Phase of sinh z.

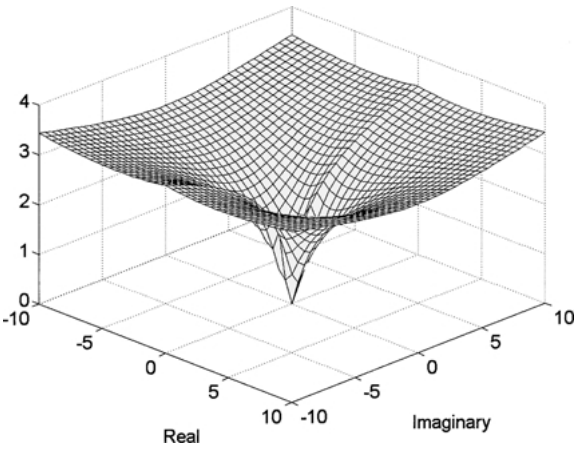


(a)

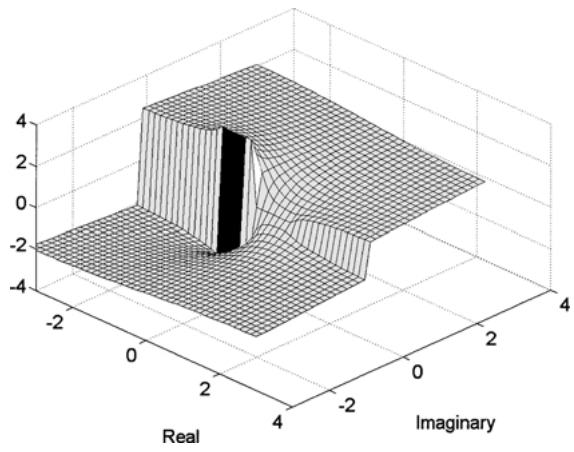


(b)

Figure 11. (a) Magnitude of arctanh z. (b) Phase of arctanh z.



(a)



(b)

Figure 12. (a) Magnitude of arcsinh z. (b) Phase of arcsinh z.

note that, unlike arcsinh z and arccos z functions that are unbounded, split-tanh x function is bounded by $\sqrt{2}$ in magnitude from the squared sum of real and imaginary tanh x components. Instead, when the domain is bounded, arcsinh z and arccos z are naturally bounded. As will be shown in numerical examples in Section 4, the radial symmetricity of arcsinh z function yields the best symbol-error-rate (SER) test result in restoring mild nonlinear distortion of non-constant modulus signals. For severe nonlinearity, arctan z and arctanh z tend to show better SER performance where the saddle-point-like magnitude response and smooth nonlinear phase response seem to provide better performance.

5. Numerical Examples

The following discrete-input and discrete-output relationship describes the TWTA AM/AM and AM/PM models using Volterra series [12]:

$$x_n = \sum_{k=1}^{\infty} \sum_{n_1} \cdots \sum_{n_{2k-1}} a_{n-n_1} a_{n-n_2} \cdots a_{n-n_k} a^*_{n-n_{k+1}} \cdots a^*_{n-n_{2k-1}} H^{(2k-1)}_{n_1,\ldots n_{2k-1}} + v_o \tag{23}$$

where v_o 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 as follows

$$
\begin{aligned}
\text{Linear coefficient:} \quad & H_0^{(1)} = 1.22 + i0.646, \\
& H_1^{(1)} = 0.063 - i0.001, \\
& H_2^{(1)} = -0.024 - i0.014, \\
& H_3^{(1)} = 0.036 + i0.031 \\
\text{3rd order coefficient:} \quad & H_{002}^{(3)} = 0.039 - i0.022, \\
& H_{330}^{(3)} = 0.018 - i0.018, \\
& H_{001}^{(3)} = 0.035 - i0.035, \\
& H_{003}^{(3)} = -0.04 - i0.009, \\
& H_{110}^{(3)} = -0.01 - i0.017 \\
\text{5th order coefficient:} \quad & H_{00011}^{(5)} = 0.039 - i0.022.
\end{aligned}
$$

Figure 13 shows the time-delayed MLP structure used for the equalization of non-constant modulus QAM signals transmitted through the channel model given in Eq. (23) to test the performance of the fully complex MLPs we introduce. The MLP equalizer used in this section has a 10-5-1 structure, i.e., a 10 tap delayed input layer fully connected to 5 neurons in the hidden layer that is connected to the single output of the equalizer. The order of the CLMS equalizer tested is 50. The sizes of equalizers are chosen so that each structure tested has similar complexity with satisfactory representation capability. A real valued learning rate is used to optimize the training in the results we present here. Tests are also carried out with imaginary and complex learning rates and it is noted that they result in some but not very significant differences in training performance. However, this is a problem that needs further study.

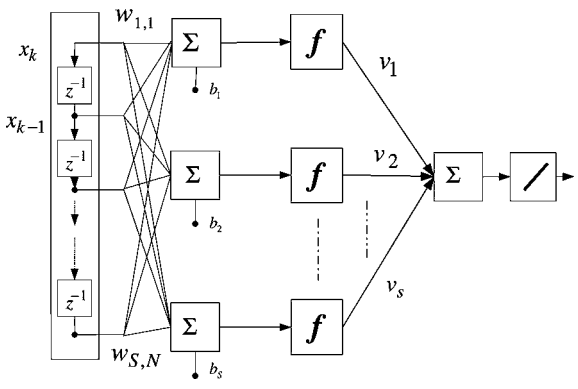


Figure 13. Nonlinear channel equalizer structure.

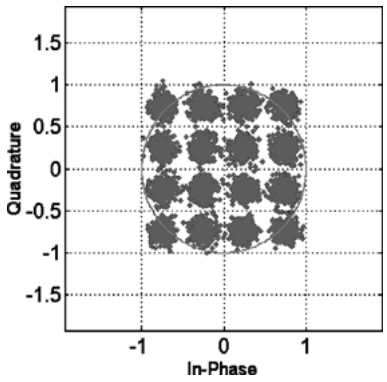


Figure 14. Input constellation plus noise.

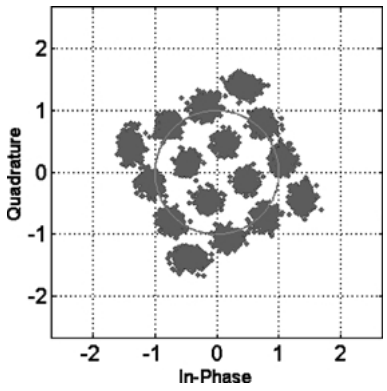


Figure 15. Received constellation with mild channel distortion.

Figure 14 shows a 16-QAM constellation with 8 dB signal-to-noise ratio per bit (Eb/No) that reaches the saturation level of unit amplitude at the outermost corner points. Figure 15 shows the AM/AM and AM/PM impact on 16-QAM symbols transmitted through TWTA where the amplitude amplification and phase rotation effects are shown. In this case, due to the moderate saturation level, the distortion is almost linear.

Figures 16 and 17 show signal-to-noise per bit (SNR/bit or Eb/No) versus SER equalization test performances of the 13 schemes discussed in this paper after each scheme is trained to restore the mild nonlinear TWTA distortion shown in Fig. 15. They include the fully complex activation functions of Georgiou and Koutsougeras (GK) [3] and Hirose [5], the split complex hyperbolic tangent, and the CLMS, along with the nine transcendental functions that were introduced in Section 4. To compare the best SER performances of all 13 schemes, we train and test all schemes at mean square error (MSE) stopping levels of 10^{-5}, 10^{-6}, and 10^{-7} at 12 dB Eb/No and pick the stopping level that yields the best overall test performance for each

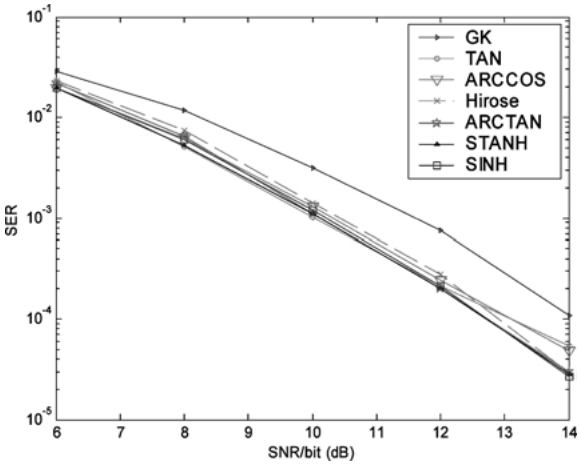


Figure 16. SER performance for the mild nonlinear distortion case (First group).

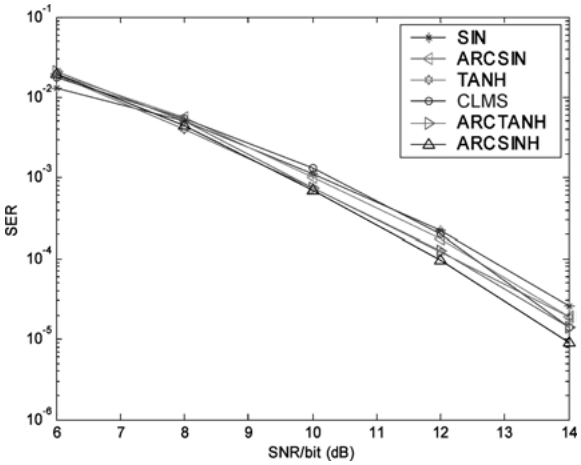


Figure 17. SER performace for the mild nonlinear distortion case (Second group).

tested scheme. Note that for most schemes, 10^{-7} MSE level provided the best performance that we had determined as the lowest threshold to be tested. In cases the schemes did not achieve the lowest MSE level of 10^{-7}, we stopped the training after 20,000 iterations. Figures 16 and 17 show these best SER performances.

Here, most schemes except that of Georgiou and Koutsougeras converged close to the minimum SER level of a 16-QAM scheme i.e., the performance with additive white gaussian noise (AWGN) only. Here, it is assumed that the noise amplitude and phase are Gaussian and uniform distributions and are independent of each other. It is not surprising that CLMS that fully utilizes the information from the real and imaginary positions performed well in this almost

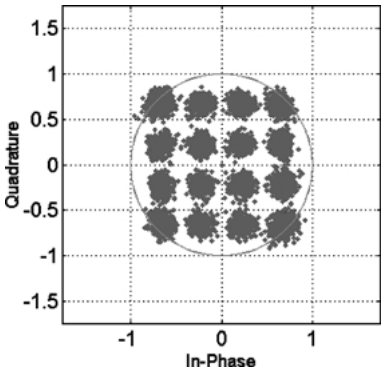


Figure 18. Equalization by arcsinh MLP (mild case).

linear distortion case. Two of the nine fully complex activation functions, arctanh z and asinh z, however, outperformed CLMS. Note that the split tanh x (STANH) did not perform well as shown in the equalization result of Fig. 19, where the overall shape of restored constellation is not as clearly square as the restored constellation of arcsinh z as shown in Fig. 18.

The second equalization example is designed to test the performance of fully complex MLPs under higher level of nonlinear distortion as illustrated in Fig. 20. The extreme level of AM/PM distortion is clearly visible here as the level of saturation is much higher for the outer constellation points. Figure 20 shows the outcome of equalization using arcsinh z activation function MLP where the equalization could not fully restore the severe AM/PM distortion from high level of saturation.

Figure 22 shows the best performance of 13 schemes in highly nonlinear saturated channel equalization case described in Figs. 20 and 21 above. In this case, the relative poor performance of CLMS along with Georgiou and Koutsougeras (GK) scheme [3] and Hirose scheme [5] are clearly represented. This is explainable in the

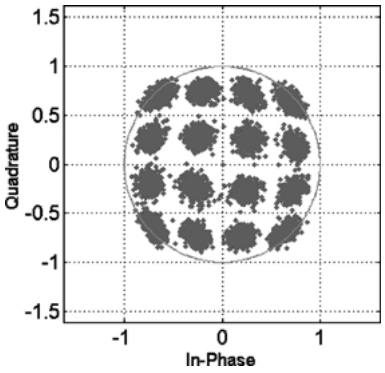


Figure 19. Equalization by STANH MLP (mild case).

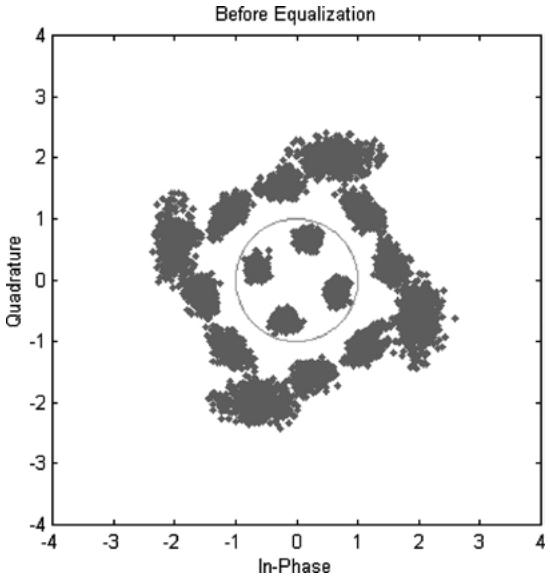


Figure 20. Received 16 QAM constelation (severe nonlinear case).

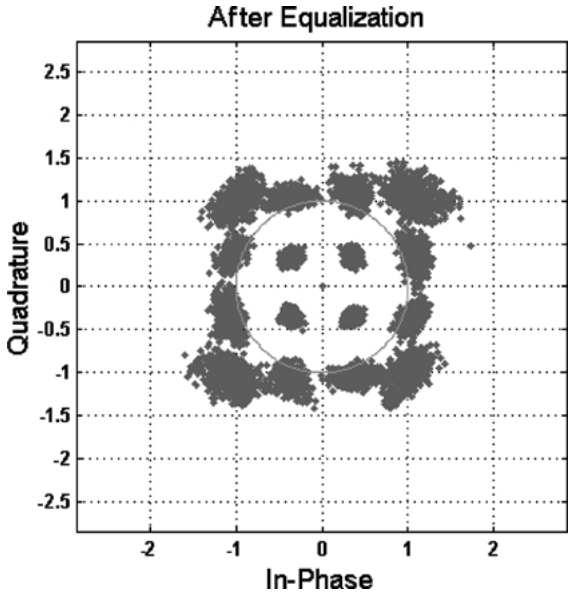


Figure 21. Equalization by arcsinh MLP (severe nonlinear case).

poor phase learning capability of these activation functions that could not restore the severe phase distortion. It is also noteworthy that unlike to the mild nonlinear case, split tanh (STANH) function outperformed CLMS as it provides nonlinear correction but four fully activation functions including $\tanh z$, $\arctan z$, $\operatorname{arctanh} z$ and $\arcsin z$ functions still outperformed the STANH function. The change of top performers between the mild and severe nonlinear distortions except $\operatorname{arctanh} z$

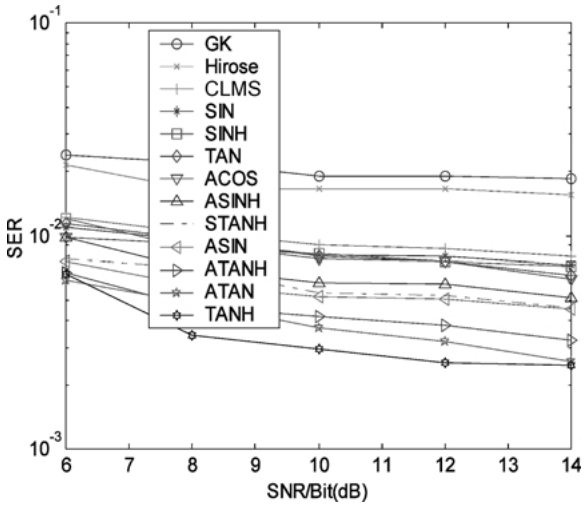


Figure 22. Eb/No vs. SER performance of severe nonlinear distortion case.

raise an intriguing question about the characteristics of activation functions under different circumstances.

6. Conclusion

Nine elementary transcendental functions are identified as fully complex nonlinear activation functions capable of learning nonlinear amplitude and phase distortion simultaneously. The fully complex MLP employing these activation functions are shown to outperform the traditional complex MLP and complex LMS schemes in satellite nonlinear channel equalization simulations. The common characteristics of elementary transcendental functions are analyzed and categorized. Most importantly, it is noted that these functions are bounded and analytic *almost everywhere* in a bounded domain of interest. This fact establishes the practicality of a fully complex MLP for its *almost everywhere* convergence property that is sufficient in all practical engineering applications.

While studying the elementary transcendental functions as the nonlinear activation function, a compact form of fully complex back-propagation algorithm is established. The *Cauchy-Riemann* equations satisfied by the *almost everywhere* analyticity of fully complex activation functions enable the simplification of the previously published complex back-propagation algorithm. It is shown that the back-propagation complex weight update equation is a complex conjugate version of the well-established real weight update equation. Furthermore, it is shown that the split complex

back-propagation algorithm is a special case of the fully complex algorithm we define. Consequently, the existing set of conditions considered to be 'desirable' for complex nonlinear activation functions from the split complex perspective [3, 4] is relaxed and reduced into a single condition.

In numerical examples, it is demonstrated that the conventional linear approach, i.e., the CLMS algorithm provides quite satisfactory performance in the case of mild nonlinear channel distortion, but performs poorly in strong nonlinear distortion. The split tanh x activation function outperforms CLMS in the strong nonlinear distortion case, but is outperformed by several fully complex elementary transcendental functions in both mild and strong nonlinear distortion cases. In every case, the singularity and discontinuity of elementary transcendental functions around the unit circle did not pose difficulty in training the fully complex MLP by backpropagation.

References

1. H. Silverman, *Complex Variables*, Houghton, Newark, USA, 1975.
2. T. Clarke, "Generalization of Neural Network to the Complex Plane," in *Proc. of IJCNN*, vol. 2, 1990, pp. 435–440.
3. G. Georgiou and C. Koutsougeras, "Complex Backpropagation," *IEEE Trans. on Circuits and Systems II*, vol. 39, no. 5, 1992, pp. 330–334.
4. C. You and D. Hong, "Nonlinear Blind Equalization Schemes Using Complex-Valued Multilayer Feedforward Neural Networks," *IEEE Trans. on Neural Networks*, vol. 9, no. 6, 1998, pp. 1442–1455.
5. D. Mandic and J. Chambers, *Recurrent Neural Networks for Prediction*, John Wiley and Sons, 2001.
6. A. Hirose, "Continuous Complex-Valued Back-Propagation Learning," *Electronics Letters*, vol. 28, no. 20, 1992, pp. 1854–1855.
7. H. Leung and S. Haykin, "The Complex Backpropagation Algorithm," *IEEE Trans. on Signal Proc*., vol. 3, no. 9, 1991, pp. 2101–2104.
8. N. Benvenuto, M. Marchesi, F. Piazza, and A. Uncini, "Non Linear Satellite Radio Links Equalized Using Blind Neural Networks," in *Proc. of ICASSP*, vol. 3, 1991, pp. 1521–1524.
9. N. Benvenuto and F. Piazza, "On the Complex Backpropagation Algorithm," *IEEE Trans. on Signal Processing*, vol. 40, no. 4, 1992, pp. 967–969.
10. M. Ibnkahla and F. Castanie, "Vector Neural Networks for Digital Satellite Communications," in *Proc. of ICC*, vol. 3, 1995, pp. 1865–1869.
11. A. Uncini, L. Vecci, P. Campolucci, and F. Piazza, "Complex-Valued Neural Networks with Adaptive Spline Activation Functions," *IEEE Trans. on Signal Processing*, vol. 47, no. 2, 1999.
12. S. Bandito and E. Biglieri, "Nonlinear Equalization of Digital Satellite Channels," *IEEE Jour. on SAC*., vol. SAC-1., 1983, pp. 57–62.
13. G. Kechriotis and E. Manolakos, "Training Fully Recurrent Neural Networks with Complex Weights," *IEEE Trans. on Circuits and Systems—II: Analog and Digital Signal Processing*, vol. 41, no. 3, 1994, pp. 235–238.
14. J. Deng, N. Sundararajan, and P. Saratchandran, "Communication Channel Equalization Using Complex-Valued Minimal Radial Basis Functions Neural Network," in *Proc. of IEEE IJCNN 2000*, vol. 5, 2000, pp. 372–377.
15. K.Y. Lee and S. Jung, "Extended Complex RBF and its Application to M-QAM in Presence of Co-Channel Interference," *Electronics Letters*, vol. 35, no. 1, 1999, pp. 17–19.
16. S. Chen, P.M. Grant, S. McLaughlin, and B. Mulgrew, "Complex-Valued Radial Basis Function Networks," in *Proc. of third IEEE International Conference on Artificial Neural Networks*," 1993, pp. 148–152.
17. T. Kim and T. Adalı, "Fully Complex Backpropagation for Constant Envelop Signal Processing," in *Proc. of IEEE Workshop on Neural Networks for Sig. Proc*., Sydney, Dec. 2000, pp. 231–240.
18. T. Kim and T. Adalı, "Complex Backpropagation Neural Network Using Elementary Transcendental Activation Functions," in *Proc. of IEEE ICASSP, Proc*. vol. II, Salt Lake City, May 2001.
19. T. Kim and T. Adalı, "Nonlinear Satellite Channel Equalization Using Fully Complex Feed-Forward Neural Networks," in *Proc. of IEEE Workshop on Nonlinear Signal and Image Processing*, Baltimore, June, 2001, pp. 141–150.

Taehwan Kim received his B.S. degree in mathematics from Seoul National University, Korea in 1984 and M.S. degree in computer science from the University of South Carolina, Columbia, U.S. in 1986. Between 1990 and 1994, he studied applied mathematics and electrical engineering as a part time graduate student at the University of Maryland College Park. Currently, he is a Ph.D. candidate in electrical engineering at the University of Maryland Baltimore County. From 1987 to 1997, he worked with Glynn Scientific, Inc., Stanford Telecommunications, Inc., and Hughes Information Systems, all in Maryland in the areas of target recognition and satellite communications. Since 1997, he has been working with the MITRE Corporation in McLean, Virginia where his main tasks have been the modernization of civil Global Positioning Satellite (GPS) and its augmentation systems for the U.S. Federal Aviation Administration (FAA). He has coauthored several papers for the Institute of Navigation (ION) conferences in the areas of robust GPS navigation, ionospheric scintillation, and interference assessment and mitigation. Since 2000, he has been presenting the mathematical property of fully complex neural networks in the IEEE International Workshop on Neural Network for Signal Processing (NNSP) and International Conference on Acoustics, Speech, and Signal

Processing (ICASSP). His research interests include adaptive signal processing, wireless communication, radio navigation, and pattern recognition.
tkim@mitre.org

Tülay Adalı received the B.S. degree from Middle East Technical University, Ankara, Turkey, in 1987 and the M.S. and Ph.D. degrees from North Carolina State University, Raleigh, in 1988 and 1992 respectively, all in electrical engineering. In 1992, she joined the Department of Electrical Engineering at the University of Maryland Baltimore County, Baltimore, where she currently is an associate professor. She has served on the organizing committees of a number of international conferences including the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), and the IEEE International Workshop on Neural Networks for Signal Processing (NNSP). She has been the general co-chair for the NNSP workshops for the last two years. She is the secretary for the IEEE Neural Networks for Signal Processing Technical Committee and is serving on the IEEE Signal Processing conference board. Her research interests are in the areas of adaptive signal processing, neural computation, estimation theory, and their applications in channel equalization, biomedical image analysis, and optical communications. Dr. Adalı is the recipient of a 1997 National Science Foundation CAREER Award.
adali@engr.umbc.edu