Neural Networks 69 (2015) 44–50

Contents lists available at ScienceDirect

Neural Networks

journal homepage: www.elsevier.com/locate/neunet

Convergence analysis of an augmented algorithm for fully complex-valued neural networks☆

Dongpo Xu [a,b,*], Huisheng Zhang [c], Danilo P. Mandic [d]

[a] *School of Mathematics and Statistics, Northeast Normal University, Changchun 130024, China*
[b] *College of Science, Harbin Engineering University, Harbin 150001, China*
[c] *Department of Mathematics, Dalian Maritime University, Dalian 116026, China*
[d] *Department of Electrical and Electronic Engineering, Imperial College London, SW7 2AZ London, UK*

ARTICLE INFO

Article history:
Received 29 June 2014
Received in revised form 28 February 2015
Accepted 17 May 2015
Available online 27 May 2015

Keywords:
Complex-valued neural networks
Augmented algorithm
Unified mean value theorem
Wirtinger calculus
Convergence

ABSTRACT

This paper presents an augmented algorithm for fully complex-valued neural network based on Wirtinger calculus, which simplifies the derivation of the algorithm and eliminates the Schwarz symmetry restriction on the activation functions. A unified mean value theorem is first established for general functions of complex variables, covering the analytic functions, non-analytic functions and real-valued functions. Based on so introduced theorem, convergence results of the augmented algorithm are obtained under mild conditions. Simulations are provided to support the analysis.

1. Introduction

Complex-valued neural networks (CVNNs) have recently attracted broad research interests, for example, in seismics, sonar, and radar (Hirose, 2012). CVNNs have been shown to inherent advantages in reducing the number of parameters and operations involved (Mandic & Goh, 2009). In addition, CVNNs have computational advantages over real-valued neural networks in solving classification problems (Aizenberg, 2011), and can even solve the XOR problem with only one complex-valued neuron (Nitta, 2003). However, the choice of activation function remains being a challenging task due to the conflict requirements of boundedness and analyticity—Liouville's theorem states that if a function is analytic and bounded in the complex plane, then it must be a constant. A traditional split-complex approach (Nitta, 1997) uses a pair of real-valued activation functions to process the real and imaginary parts of a complex signal separately. While this approach helps avoiding the problem of unboundedness, split complex activation functions are never analytic. In contrast, 'fully' complex activation functions (Kim & Adali, 2003), such as elementary transcendental functions, are analytic and bounded almost everywhere in $\mathbb{C}$, and have been used in multi-layer perceptions (Kim & Adali, 2003), radial basis function networks (Savitha, Suresh, & Sundararajan, 2012) and extreme learning machines (Li, Huang, Saratchandran, & Sundararajan, 2005). Classical real-valued learning algorithms that have been extended to complex case, contains the complex least mean square (Widrow, McCool, & Ball, 1975), complex backpropagation (Georgiou & Koutsougeras, 1992; Hirose, 1992; Leung & Haykin, 1991; Nitta, 1997) and complex real-time recurrent learning (Goh & Mandic, 2004, 2007a). Signal processing techniques (Adali, Li, Novey, & Cardoso, 2008; Dini & Mandic, 2012) have also been proposed based on such activation functions, however, the basic research issue: whether these complex algorithms share convergence properties with their real counterparts remains largely unanswered. The complex universal approximation theorem of the CVNNs with fully complex activation functions (denoted as FCVNNs for simplicity) has been given by Kim and Adali in Kim and Adali (2003), which ensures that the FCVNNs can be considered as a universal approximator of any continuous complex mappings.

Convergence of the real-valued learning algorithm has been widely studied (Shao & Zheng, 2011; Wang, Yang, & Wu, 2011; Wu, Fan, & Zurada, 2014; Wu, Feng, Li, & Xu, 2005; Wu, Wang, Cheng, & Li, 2011). However, in the complex domain, in addition to

☆ This work was supported by the National Natural Science Foundation of China (Nos. 61301202, 61101228) and by the Research Fund for the Doctoral Program of Higher Education of China (No. 20122304120028).

* Corresponding author at: School of Mathematics and Statistics, Northeast Normal University, Changchun 130024, China.

E-mail addresses: dongpoxu@gmail.com (D. Xu), zhhuisheng@163.com (H. Zhang), d.mandic@imperial.ac.uk (D.P. Mandic).

http://dx.doi.org/10.1016/j.neunet.2015.05.003

the conflict between boundedness and analyticity of the activation function, another challenge is that the traditional mean value theorem does not hold in the complex domain (e.g., $f(z) = e^z$ with $z_2 = z_1 + 2\pi i$, then $f(z_1) = f(z_2)$ but $f(z_2) - f(z_1) \neq f'(\xi)(z_2 - z_1)$ for all $\xi \in \mathbb{C}$). In addition, cost functions are real-valued and therefore the complex derivative cannot be used. Some results for split-complex nonlinear gradient descent (SCNGD) algorithms exist (Xu, Zhang, & Liu, 2010; Zhang, Zhang, & Wu, 2009), whereby the analysis is based on reformulating complex algorithm in the real domain by separating it into real and imaginary parts. Furthermore, the convergence of the SCNGD algorithms with momentum or penalty has been established in Xu, Shao, and Zhang (2012) and Zhang, Xu, and Zhang (2014). In addition, the convergence of some complex adaptive filters algorithms has been obtained under the assumption that the activation function is a contraction (Mandic & Goh, 2009). The convergence of fully-complex nonlinear gradient descent (FCNGD) algorithms has been proved under Schwarz symmetry condition $f^*(z) = f(z^*)$ (Zhang, Liu, Xu, & Zhang, 2014). However, this condition is usually not valid for a polynomial function with complex coefficients, and the mean value theorem used in Zhang, Liu et al. (2014) is not applicable to the non-analytic functions, such as real-valued cost functions. Recently, augmented complex statistics have been introduced into some learning algorithms, such as the augmented complex least mean square (Mandic & Goh, 2009; Mandic, Javidi, Goh, Kuh, & Aihara, 2009), augmented complex extended Kalman filter (Dini & Mandic, 2012; Goh & Mandic, 2007b), and augmented echo state network (Xia, Jelfs, Van Hulle, Príncipe, & Mandic, 2011). These can capture the second-order statistical information and thus produce optimal estimates for second-order noncircular (improper) signals. However, the convergence of the augmented FCNGD (AFCNGD) algorithms for the FCVNNs has not yet been fully established in the literature, which motivates this work.

The aim of this paper is to present a comprehensive study on the weak and strong convergence for the AFCNGD algorithm, indicating that the gradient of the error function goes to zero and the weight sequence goes to a fixed point, respectively. In comparison to the existing complex backpropagation (CBP) algorithms (Georgiou & Koutsougeras, 1992; Hirose, 1992; Leung & Haykin, 1991; Nitta, 1997), the proposed AFCNGD algorithm shows faster convergence and better steady-state performance. The main points and novel contributions of this paper are as follows:

- Based on Wirtinger calculus, we develop an augmented FCNGD algorithm for CVNNs with fully complex activation functions. This approach can simplify the derivation of the proposed algorithm and eliminate Schwarz symmetry restriction on the complex activation functions.
- We establish a unified mean value theorem for the complex nonlinear functions, covering the analytic functions, non-analytic functions and real-valued functions. This theorem plays an important role in the convergence proof of the proposed AFCNGD algorithm.
- The deterministic convergence including weak convergence and strong convergence of the AFCNGD algorithm is obtained. Our results are of considerable generality, including as particular cases almost all CVNNs with complex elementary transcendental functions given in Kim and Adali (2003).
- Illustrated experiments have been performed to verify the theoretical results of this paper and the advantages of the proposed AFCNGD algorithm.

The rest of this paper is organized as follows. In Section 2, we provide an overview of second-order augmented complex statistics and Wirtinger calculus. Section 3 derives the proposed augmented learning algorithm for the FCVNNs. The main convergence results and their proofs are presented in Section 4. Supporting numerical experiments are presented in Section 5. Some conclusions are drawn in Section 6.

2. Preliminaries

2.1. Notations

We use bold-face upper case letter to denote matrices, bold-faced lower case letters for column vectors, and light-faced lower case letters for scalars. The superscripts $(\cdot)^*$, $(\cdot)^T$ and $(\cdot)^H$ denote the complex conjugate, transpose and Hermitian (conjugate transpose), respectively. $\mathrm{Re}(z)$ and $|z|$ denote the real part and module of a complex number z. $\|\mathbf{z}\|$ and $\|\mathbf{Z}\|$ denote the Frobenius norm of a vector $\mathbf{z}$ and a matrix $\mathbf{Z}$. Finally, we refer to $f(z^*) = f^*(z)$ as the Schwarz symmetry principle (Needham, 1998, p. 257).

2.2. Second-order augmented complex statistics

The recent introduction of so-called augmented complex statistics (Mandic & Goh, 2009) showed that for a general (improper) complex vector $\mathbf{z}$, second order statistics based on the covariance matrix $\mathbf{C_{zz}} = E[\mathbf{z}\mathbf{z}^H]$ is inadequate and that the pseudo-covariance matrix $\mathbf{P_{zz}} = E[\mathbf{z}\mathbf{z}^T]$ is also required to fully capture the second order information. Processes with the vanishing pseudo-covariance, $\mathbf{P_{zz}} = \mathbf{0}$ is termed second order circular (or proper). In real-world applications, most complex signals are second order noncircular or improper, and their probability density functions are not rotation invariant. In practice, the widely linear modeling (Picinbono & Chevalier, 1995) is based on a regressor vector produced by concatenating the input vector $\mathbf{z}$ with its complex conjugate $\mathbf{z}^*$, to give an augmented $2M \times 1$ input vector $\mathbf{z}^a = [\mathbf{z}^T, \mathbf{z}^H]^T$, together with the corresponding augmented coefficient vector $\mathbf{w}^a = [\mathbf{u}^T, \mathbf{v}^H]^T$. The $2M \times 2M$ augmented covariance matrix (Schreier & Scharf, 2003) then becomes

$$\mathbf{C}_{\mathbf{z}^a\mathbf{z}^a} = E\begin{bmatrix}\mathbf{z}\\ \mathbf{z}^*\end{bmatrix}\begin{bmatrix}\mathbf{z}^H & \mathbf{z}^T\end{bmatrix} = \begin{pmatrix}\mathbf{C_{zz}} & \mathbf{P_{zz}}\\ \mathbf{P}^*_{\mathbf{zz}} & \mathbf{C}^*_{\mathbf{zz}}\end{pmatrix}. \tag{1}$$

This matrix now contains the complete complex second order statistical information available in the complex domain, see Mandic and Goh (2009) and Schreier and Scharf (2010) for more details.

2.3. Wirtinger calculus

Any function of a complex variable z can be defined as $f(z) = u(x, y) + iv(x, y)$, where $z = x + iy$ and i denotes an imaginary unit. If the partial derivatives $\frac{\partial u}{\partial y}$, $\frac{\partial v}{\partial x}$, $\frac{\partial u}{\partial x}$, $\frac{\partial v}{\partial y}$ exist and satisfy the Cauchy–Riemann conditions $\frac{\partial u}{\partial x} = \frac{\partial v}{\partial y}$ and $\frac{\partial v}{\partial x} = -\frac{\partial u}{\partial y}$, then $f(z)$ is said to be analytic (complex derivative exists), otherwise, it is non-analytic (complex derivative does not exist). For general functions of complex variables (both analytic and non-analytic), the following pair of derivatives can be defined (Brandwood, 1983; Kreutz-Delgado, 2009; Wirtinger, 1927)

$$\text{R-derivative: } \frac{\partial f}{\partial z} = \frac{1}{2}\left(\frac{\partial f}{\partial x} - i\frac{\partial f}{\partial y}\right) \tag{2}$$

$$\text{R}^*\text{-derivative: } \frac{\partial f}{\partial z^*} = \frac{1}{2}\left(\frac{\partial f}{\partial x} + i\frac{\partial f}{\partial y}\right) \tag{3}$$

which are called Wirtinger or CR derivatives. In particular, if $f(z)$ is analytic, then the R-derivative $\frac{\partial f}{\partial z}$ becomes the complex derivative $f'(z)$ and the R*-derivative vanishes, that is the Cauchy–Riemann equations are equivalent to $\frac{\partial f}{\partial z^*} = 0$. Some basic rules of the CR derivatives are summarized as (Kreutz-Delgado, 2009; Mandic & Goh, 2009; Wirtinger, 1927)

$$\text{Differential rule: } df = \frac{\partial f}{\partial z}dz + \frac{\partial f}{\partial z^*}dz^* \tag{4}$$