PERGAMON

AVAILABLE AT
www.ComputerScienceWeb.com
POWERED BY SCIENCE DIRECT®

Neural Networks 16 (2003) 155–159

Neural Networks

www.elsevier.com/locate/neunet

Neural Networks Letter

A complex-valued nonlinear neural adaptive filter with a gradient adaptive amplitude of the activation function

Andrew I. Hanna[a,*], Danilo P. Mandic[b,1]

[a]School of Information Systems, University of East Anglia, Norwich, NR4 7TJ Norfolk, UK
[b]Communications and Signal Processing Group, Department of Electrical and Electronic Engineering, Imperial College of Science, Technology and Medicine, London, UK

Received 30 January 2002; accepted 24 October 2002

Abstract

A complex-valued nonlinear gradient descent (CNGD) learning algorithm for a simple finite impulse response (FIR) nonlinear neural adaptive filter with an adaptive amplitude of the complex activation function is proposed. This way the amplitude of the complex-valued analytic nonlinear activation function of a neuron in the learning algorithm is made gradient adaptive to give the complex-valued adaptive amplitude nonlinear gradient descent (CAANGD). Such an algorithm is beneficial when dealing with signals that have rich dynamical behavior. Simulations on the prediction of complex-valued coloured and nonlinear input signals show the gradient adaptive amplitude, CAANGD, outperforming the standard CNGD algorithm.

Keywords: Complex-valued nonlinear gradient descent; Complex-valued adaptive amplitude nonlinear gradient descent; Finite impulse response filter

1. Introduction

Recently, there has been an increased interest in the processing of complex-valued nonlinear signals (Hirose 1992; Georgiou & Koutsougeras, 1992). To this cause, the class of least mean square (LMS) based adaptive finite impulse response (FIR) filters for real-valued signals have been extended to the complex plane. The complex-valued LMS (CLMS) algorithm, (Widrow, McCool, & Ball, 1975), was the fundamental foundation for the class of complex adaptive filters that followed. It is well known that for the family of complex-valued gradient descent algorithms to perform, the nonlinearity in the activation function must be analytic and bounded almost everywhere in the complex plane, $\mathbb{C}$. The complex nonlinear gradient descent (CNGD) and the complex backpropagation (CBP) algorithms, (Georgiou & Koutsougeras, 1992; Benvenuto & Piazza, 1992; Hanna & Mandic, 2002), employ a complex nonlinear activation function to extend the class of linear complex adaptive filters to nonlinear complex signals. The aim of this paper is to extend the derivation of the gradient descent based adaptive amplitude of the activation function for real-valued neural networks as introduced in Trentin (2001) to the case of complex-valued nonlinear activation functions. Hence, for convenience, we consider a simple complex-valued dynamical feedforward perceptron employed as a nonlinear FIR adaptive filter, shown in Fig. 1. This architecture operates for both real and complex-valued neural adaptive filters. By making the amplitude of the activation function gradient adaptive, such a nonlinear adaptive filter shows better performance on signals with rich dynamics than the standard learning algorithm.

2. The complex-valued nonlinear gradient descent (CNGD) algorithm

The complex-valued gradient descent algorithms for a complex-valued nonlinear perceptron employed as a

* Corresponding author. Tel.: +44-1603-592-442; fax: +44-1603-593-345.
E-mail addresses: aih@sys.uea.ac.uk (A.I. Hanna), d.mandic@ic.ac.uk (D.P. Mandic).
[1] Tel.: +44-207-594-6271; fax: +44-207-594-6234.

0893-6080/03/$ - see front matter
PII: S0893-6080(02)00236-8

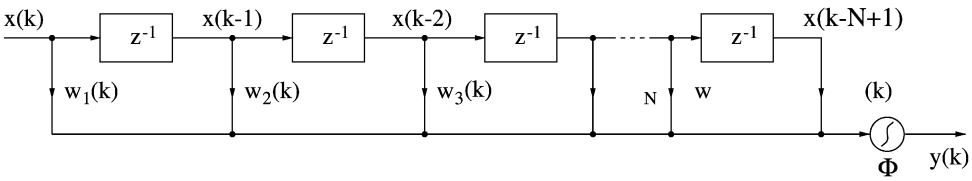


Fig. 1. A complex-valued nonlinear FIR filter.

nonlinear FIR filter are based upon an objective function given by

$$J(k) = \frac{1}{2}[e(k)e^*(k)] = \frac{1}{2}|e(k)^2|, \qquad (1)$$

$$e(k) = d(k) - \Phi(\mathbf{x}^{\mathrm{T}}(k)\mathbf{w}(k)),$$

where $e(k)$ denotes the instantaneous output error from the filter, $d(k)$ the desired response, $\mathbf{x}(k) = [x_1(k), \ldots, x_N(k)]^{\mathrm{T}}$ the input to the filter, $\mathbf{w}(k) = [w_1(k), \ldots, w_N(k)]^{\mathrm{T}}$ the weight vector and $\Phi(\cdot)$ some complex-valued nonlinearity within the neuron. The superscripts $(\cdot)^{\mathrm{T}}$ and $(\cdot)^*$ denote the transpose and complex conjugate operators, whereas the superscripts $(\cdot)^{\mathrm{r}}$ and $(\cdot)^{\mathrm{i}}$ denote the real and imaginary parts, respectively, and $\mathrm{j} = \sqrt{-1}$. For simplicity, we shall denote $\Phi(\mathbf{x}^{\mathrm{T}}(k)\mathbf{w}(k)) = \Phi(k) = u(k) + \mathrm{j}v(k)$, and thus

$$e^{\mathrm{r}}(k) = d^{\mathrm{r}}(k) - u(k), \qquad e^{\mathrm{i}}(k) = d^{\mathrm{i}}(k) - v(k). \qquad (2)$$

The learning algorithm in this case is the CNGD algorithm, defined by (Mandic & Chambers, 2001)

$$\mathbf{w}(k+1) = \mathbf{w}(k) + \Delta\mathbf{w}(k), \qquad (3)$$

$$\Delta\mathbf{w}(k) = \eta e(k)[\Phi'(\mathbf{x}^{\mathrm{T}}(k)\mathbf{w}(k))]^*\mathbf{x}^*(k), \qquad (4)$$

where η denotes the step size of the algorithm.

3. The complex-valued adaptive amplitude nonlinear gradient descent algorithm

The complex-valued adaptive amplitude nonlinear gradient descent (CAANGD) algorithm is an extension to the CNGD algorithm. The CAANGD relies on the amplitude of the analytic nonlinear activation function to be adaptive according to the change in dynamics of the input signal. We can extend the activation function with range λ from Trentin (2001) as

$$\Phi(k) = \lambda\bar{\Phi}(k) = \lambda(\bar{u}(k) + \mathrm{j}\bar{v}(k)), \qquad (5)$$

where λ denotes the amplitude of the nonlinearity, $\Phi(k)$, whereas $\bar{\Phi}(k)$ denotes the activation function with unit amplitude. For the logistic sigmoid function this would be

$$\Phi(x(k), \beta, \lambda) = \frac{\lambda}{1 + \mathrm{e}^{-\beta x(k)}},$$

where $x(k) \in \mathbb{C}$. Thus if $\lambda = 1$ it follows that $\Phi(k) = \bar{\Phi}(k)$. Here, we propose a gradient adaptive amplitude of the analytic nonlinear activation function in order to increase performance of the nonlinear complex-valued filter. The update for the gradient adaptive amplitude is given by, (Trentin, 2001)

$$\lambda(k+1) = \lambda(k) - \rho\nabla_\lambda J(k)_{|\lambda=\lambda(k)}, \qquad (6)$$

where $\nabla_\lambda J(k)_{|\lambda=\lambda(k)}$ denotes the gradient of the objective function (1), with respect to the amplitude of the activation function λ, and $\rho \in \mathbb{R}$ denotes the step size of the algorithm and is chosen to be a small constant. Having this in mind we can then deal with a general complex-valued structure (Mandic & Chambers, 2001), it can then be shown that

$$\nabla_\lambda J(k)_{|\lambda=\lambda(k)} = \frac{\partial J(k)}{\partial\lambda(k)} = \frac{1}{2}\frac{\partial[e(k)e^*(k)]}{\partial\lambda(k)} = \frac{1}{2}\left[e^*(k)\frac{\partial e(k)}{\partial\lambda(k)} + e(k)\frac{\partial e^*(k)}{\partial\lambda(k)}\right], \qquad (7)$$

stating that

$$\frac{\partial e^*(k)}{\partial\lambda(k)} = \left(\frac{\partial e(k)}{\partial\lambda(k)}\right)^*,$$

notice that λ is real-valued, therefore the derivative can be expanded to give,

$$\frac{\partial e(k)}{\partial\lambda(k)} = \frac{\partial e^{\mathrm{r}}(k)}{\partial\lambda(k)} + \mathrm{j}\frac{\partial e^{\mathrm{i}}(k)}{\partial\lambda(k)} = \frac{\partial[d^{\mathrm{r}}(k) - \lambda(k)\bar{u}(k)]}{\partial\lambda(k)} + \mathrm{j}\frac{\partial[d^{\mathrm{i}}(k) - \lambda(k)\bar{v}(k)]}{\partial\lambda(k)} = -\bar{u}(k) - \mathrm{j}\bar{v}(k) = -\bar{\Phi}(k). \qquad (8)$$

Therefore, the desired learning algorithm for complex-valued nonlinear adaptive filters becomes

$$e(k) = d(k) - \Phi(\mathbf{x}^{\mathrm{T}}(k)\mathbf{w}(k)), \quad \Phi(\mathbf{x}^{\mathrm{T}}(k)\mathbf{w}(k)) = \lambda(k)\bar{\Phi}(\mathbf{x}^{\mathrm{T}}(k)\mathbf{w}(k)), \qquad (9)$$

$$\mathbf{w}(k+1) = \mathbf{w}(k) + \eta e(k)[\Phi'(\mathbf{x}^{\mathrm{T}}(k)\mathbf{w}(k))]^*\mathbf{x}^*(k), \qquad (10)$$

$$\lambda(k+1) = \lambda(k) + \frac{\rho}{2}[e^*(k)\bar{\Phi}(\mathbf{x}^{\mathrm{T}}(k)\mathbf{w}(k)) + e(k)(\bar{\Phi}(\mathbf{x}^{\mathrm{T}}(k)\mathbf{w}(k)))^*], \qquad (11)$$

which describes the CAANGD algorithm for complex-valued feedforward dynamical perceptrons, employed as nonlinear adaptive FIR filters.

4. Experiments

To investigate the performance of the proposed algorithm compared to other algorithms of this kind, they were applied to the problem of time-series prediction, by averaging the performance curves of 100 independent simulations. For rigour, all the algorithms were tested on complex-valued coloured and nonlinear inputs. The nonlinear filter was given by (Narendra & Parthasarathy, 1990)

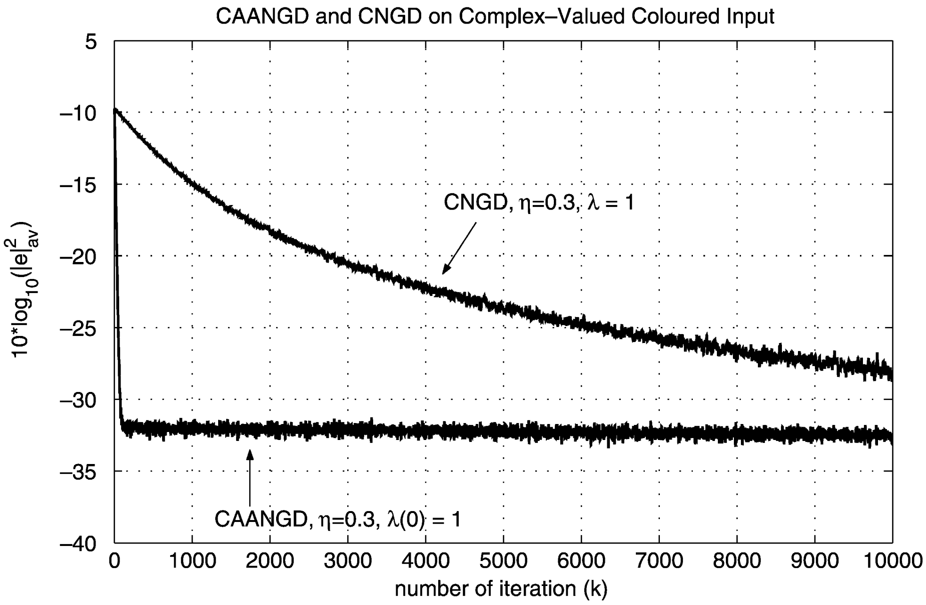
$$z(k) = \frac{z(k-1)}{1+z^2(k-1)} + r^2(k), \tag{12}$$

where $r(k)$ was a normally distributed $\mathcal{N}(0,1)$ complex-valued white noise, $n(k)$, passed through a stable AR filter given by

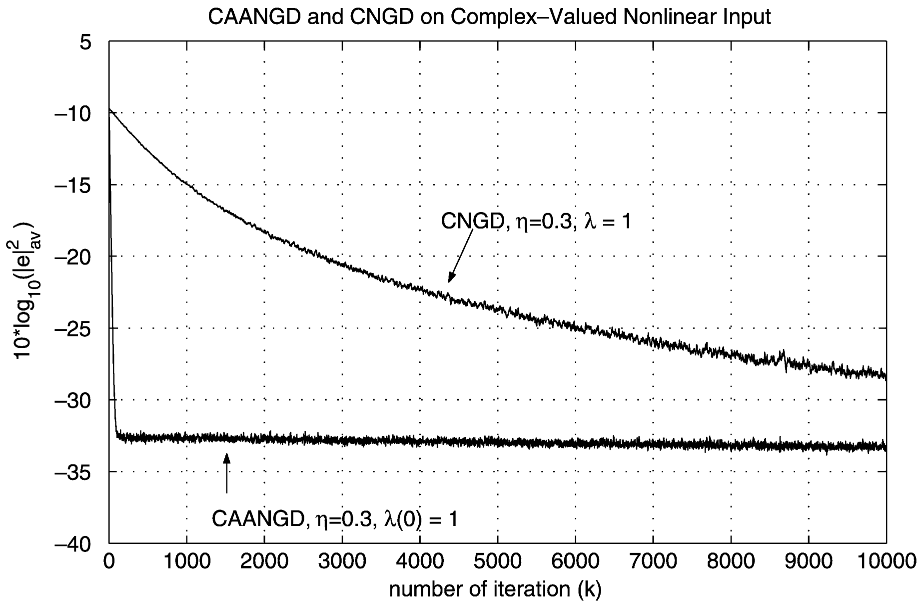
$$r(k) = 1.55r(k-1) - 0.81r(k-2) + n(k). \tag{13}$$

The input for both signals was scaled to range between $[0, 0.1]$ and the nonlinearity at the neuron was chosen to be the complex logistic sigmoid function,

$$\Phi(\mathbf{x}(k), \mathbf{w}(k), \beta, \lambda(k)) = \frac{\lambda(k)}{1+e^{-\beta\mathbf{x}^T(k)\mathbf{w}(k)}}, \tag{14}$$



(a) Performance curves for CNGD and CAANGD on complex coloured input

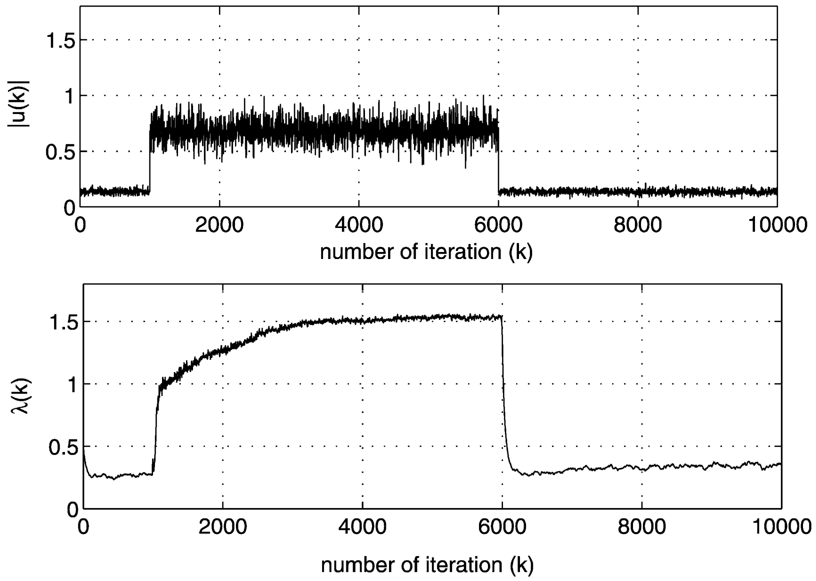


(b) Performance curves for CNGD and CAANGD on complex nonlinear input

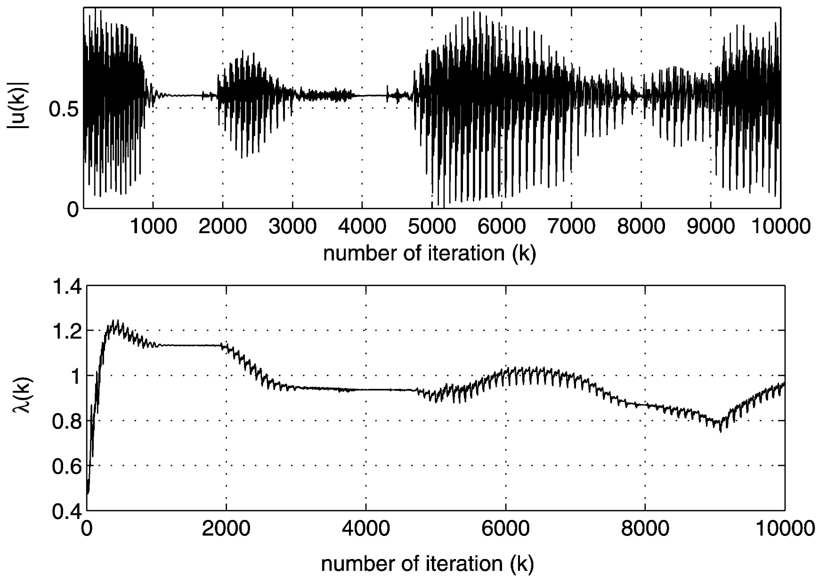
Fig. 2. Performance curves for CNGD and CAANGD on complex input.

with a slope, $\beta = 1$, learning rate $\eta = 0.3$ and an initial amplitude of $\lambda(0) = 1$. For the CAANGD algorithm, the step size of the adaptive amplitude learning algorithm was $\rho = 0.15$. Fig. 2(a) and (b) shows the performance curves for the CNGD algorithm and the CAANGD algorithm on coloured input (13) and nonlinear input (12). Both figures clearly show the CAANGD converging significantly faster than the standard CNGD algorithm, thus demonstrating the increased performance on signals with rich dynamical range.

To further investigate the algorithm, the CAANGD algorithm was used to predict signals with rich dynamics, such as speech, which for this experiment was made complex. The top diagram in Fig. 3(a) shows the magnitude of a synthetic complex-valued nonlinear input, whereas the bottom diagram shows the plot of the adaptive amplitude. It can be clearly seen that as the amplitude of the input signal increases around sample number 1000 the CAANGD adapts the amplitude of the nonlinearity accordingly. Similarly as the amplitude of the input signal is reduced, the amplitude, λ, of the nonlinearity, $\Phi(\mathbf{x}^{\mathrm{T}}(k)\mathbf{w}(k))$, adjusts accordingly. The top diagram in Fig. 3(b) shows the magnitude, $|u(k)|$, of the complex speech signal and the bottom diagram shows the plot of the adapted amplitude, λ. As in Fig. 3(a), as the amplitude of the input speech signal changes, so does the amplitude of



(a) Adaptive amplitude for prediction of complex nonlinear input



(b) Adaptive amplitude for prediction of complex speech input

Fig. 3. Adaptive amplitudes for CAANGD on complex nonlinear and speech input.

the CAANGD as it adapts to the changes in the dynamics of the input signal.

5. Conclusions

The amplitude of the activation function in the CNGD algorithm for a simple complex-valued nonlinear neural adaptive filter has been made adaptive using a gradient descent based approach to give the CAANGD algorithm. The algorithm has been developed for a general complex nonlinear activation function of a filter. The proposed algorithm has been shown to converge faster than the standard CNGD algorithm for nonlinear prediction of signals with large dynamics. The average of a series of independent simulations show the CAANGD algorithm outperforming the CNGD algorithm on complex-valued coloured and nonlinear input. In addition, experimental results show the amplitude update following the change in the dynamics of the input signal.

References

Benvenuto, N., & Piazza, F. (1992). On the complex backpropagation algorithm. *IEEE Transactions on Signal Processing*, *40*(4), 967–969.

Georgiou, G. M., & Koutsougeras, C. (1992). Complex domain backpropagation. *IEEE Transactions on Circuits and Systems-II: Analog and Digital Signal Processing*, *39*(5), 330–334.

Hanna, A. I. & Mandic, D. P (2002). A normalised complex backpropagation algorithm. *Proceedings of IEEE ICASSP'02, Orlando* (pp. 977–980).

Hirose, A (1992). Proposal of fully complex-valued neural networks. *Proceedings of IV JCNN'92, Baltimore* (pp. 152–157).

Mandic, D. P., & Chambers, J. A. (2001). *Recurrent neural networks for prediction*, London: Wiley.

Narendra, K. S., & Parthasarathy, K. (1990). Identification and control of dynamical systems using neural networks. *IEEE Transactions on Neural Networks*, *1*(1), 4–27.

Trentin, E. (2001). Networks with trainable amplitude of activation functions. *Neural Networks*, *14*(4/5), 471–493.

Widrow, B., McCool, J., & Ball, M. (1975). The complex LMS algorithm. *Proceedings of the IEEE*, *63*, 719–720.