ANALYSIS OF THE CLASS OF COMPLEX-VALUED ERROR ADAPTIVE NORMALISED NONLINEAR GRADIENT DESCENT ALGORITHMS

Andrew I. Hanna and Ian Yates

School of Information Systems,
University of East Anglia,
Norwich, UK.
{aih, iy}@sys.uea.ac.uk

Danilo P. Mandic

Dept. of Electrical Engineering
Imperial College
London, UK.
d.mandic@ic.ac.uk

ABSTRACT

A complex-valued gradient based algorithm for training nonlinear complex-valued finite impulse response (FIR) filters is derived. The proposed complex error-adaptive normalised nonlinear gradient descent (CEANNGD) and smoothed CEANNGD (SCEANNGD) algorithms are an improvement on the complex nonlinear gradient descent (CNGD) and the complex normalised nonlinear gradient descent (CNNGD) algorithm by including an adaptive term in the normalised learning rate of the CNNGD. This is achieved by performing a minimisation of the complex-valued instantaneous output error that has been approximated via a Taylor series expansion, which makes it suitable for the processing of nonlinear and nonstationary signals. Experiments on complex-valued coloured and nonlinear signals show that the CEANNGD and SCEANNGD algorithms outperform the standard CNNGD and CNGD algorithms.

1. INTRODUCTION

The data that is processed in many signal processing areas is resident in the field of real numbers $\mathbb{R}$, recently however progress in biomedicine, communications and signal processing has brought a new class of signals and it has been necessary to extend known techniques of data processing to include the field of complex numbers $\mathbb{C}$. The development of complex-valued adaptive filtering began with the development of the complex least mean square (CLMS) algorithm [1] for linear complex-valued FIR filters. As with the development of real-valued adaptive filters, the need to extend the class of algorithms from the linear to non-linear case for complex-valued adaptive filtering algorithms was apparent. The issue of finding a suitable activation function is still under investigation. Liouville's theorem that states *"a bounded entire function in the complex plane is a constant"*, and therefore a general holomorphic activation function that is bounded almost everywhere in $\mathbb{C}$ is a good choice for the nonlinearity. The effects of different activation functions used in any complex-valued nonlinear gradient descent algorithm are shown in [2]. This paper first derives the complex-valued normalised nonlinear gradient descent (CNNGD) algorithm for a dynamical perceptron shown in Figure 1. The CNNGD algorithm is derived according to the usual independence assumptions. Therefore the dynamics of the input signal are not fully considered in the derivation of the adaptive learning rate, $\eta(k)$. To this cause the paper then presents a complex-valued error-adaptive normalised nonlinear gradient descent (CEANNGD) algorithm, that minimises the instantaneous output error according to an estimate of the remainder of the Taylor series expansion used in the derivation of the CNNGD algorithm. The proposed algorithm applies to any activation function thus complementing the analysis shown in [2].

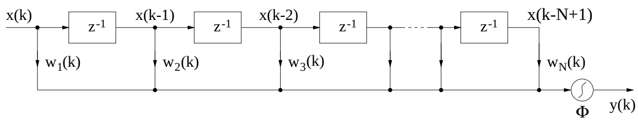

Fig. 1. Complex-valued nonlinear adaptive FIR filter

2. THE COMPLEX NONLINEAR GRADIENT DESCENT ALGORITHM

The equations that describe the complex nonlinear gradient descent (CNGD) algorithm for a complex-valued nonlinear adaptive filter are given by

$$E(k) = \frac{1}{2}|e(k)|^2 = \frac{1}{2}[e(k)e^*(k)] \tag{1}$$

where $E(k)$ is the cost function, $(\cdot)^*$ denotes the complex conjugate, and $e(k)$ is the complex instantaneous output error given by

$$e(k) = d(k) - y(k), \quad y(k) = \Phi(net(k)) \tag{2}$$

$$net(k) = \sum_{n=1}^{N} x_n(k) w_n(k) = \mathbf{x}^T(k)\mathbf{w}(k), \qquad (3)$$

where $d(k)$ is the desired output, the complex-valued input signal is given by $\mathbf{x}(k) \triangleq [x_1(k), \ldots, x_N(k)]^T$, $\mathbf{w}(k) \triangleq [w_1(k), \ldots, w_N(k)]^T$ the complex weight vector, N is the number of tap inputs, and $\Phi(net(k))$ is some nonlinear holomorphic function that is bounded almost everywhere in the complex domain. For simplicity we state that

$$\begin{aligned} \Phi(net(k)) &= \Phi^r(net(k)) + j\Phi^i(net(k)) \\ &= u(k) + jv(k), \end{aligned} \qquad (4)$$

where $j = \sqrt{-1}$ and the superscripts $(\cdot)^r$ and $(\cdot)^i$ denote the real and imaginary parts respectively. We can then split up the error term (2) into its real and imaginary parts

$$e^r(k) = d^r(k) - u(k), \qquad e^i(k) = d^i(k) - v(k). \qquad (5)$$

The individual weight adaptation in the complex nonlinear gradient descent algorithm can be written in the compact form, and is given by [3]

$$\mathbf{w}(k+1) = \mathbf{w}(k) + \eta e(k)(\Phi'[net(k)])^* \mathbf{x}^*(k). \qquad (6)$$

where η denotes the step size of the algorithm.

3. A COMPLEX NORMALISED NONLINEAR GRADIENT DESCENT ALGORITHM

Let us find an optimal learning rate that minimises the output error by expanding the error term (2) by a Taylor series expansion [4]. Assuming the usual statistical independence between weights, input signal, error, and learning rate η, the expanded error term then becomes

$$\begin{aligned} e(k+1) &= e(k) + \sum_{n=1}^{N} \frac{\partial e(k)}{\partial w_n(k)} \Delta w_n(k) \\ &+ \frac{1}{2!} \sum_{n=1}^{N} \sum_{m=1}^{N} \frac{\partial e^2(k)}{\partial w_n(k) \partial w_m(k)} \Delta w_n(k) \Delta w_m(k) + \cdots \end{aligned} \qquad (7)$$

The weight update term $\Delta w_n(k)$ has already been derived in (6), thus we must find $\frac{\partial e(k)}{\partial w_n(k)}$. Since $e(k)$ is an analytic complex function we can apply the Cauchy-Riemann equations to give [3]

$$\begin{aligned} \frac{\partial e(k)}{\partial w_n(k)} &= \frac{\partial e^r(k)}{\partial w_n^r(k)} + j \frac{\partial e^i(k)}{\partial w_n^r(k)} \\ &= -\Phi(net(k)) x_n(k). \end{aligned} \qquad (8)$$

The weight update term has already been derived in (6), therefore truncating (7) to include the first two terms, and substituting (6) and (8), gives

$$e(k+1) = e(k)[1 - \eta |\Phi'[net(k)]|^2 ||\mathbf{x}(k)||_2^2]. \qquad (9)$$

In order for the error in (9) to be zero, the term in the square brackets must be zero and can thus be used to find $\eta(k)$ as

$$\eta(k) = \frac{1}{|\Phi'[net(k)]|^2 ||\mathbf{x}(k)||_2^2 + C} \qquad (10)$$

The optimal learning rate for a complex-valued feed-forward FIR filter employed as a nonlinear adaptive filter is given in (10) and gives rise to the complex normalised nonlinear gradient descent (CNNGD) algorithm.

4. A COMPLEX-VALUED ERROR-ADAPTIVE NORMALISED NONLINEAR GRADIENT DESCENT ALGORITHM

Notice that in the derivation of the normalised learning rate $\eta(k)$ of the complex-valued normalised nonlinear gradient descent (CNNGD) algorithm, a small positive constant term C has been added to balance the truncation of the Taylor series expansion, and to prevent the learning rate from tending to infinity. Namely, the CNNGD algorithm experiences problems similar to those of the normalised least-mean square (NLMS) algorithm, which are due to the assumptions in its derivation. Therefore a constant term, C, in the CNNGD algorithm accounts for the correlation between the signal, error, and learning rate, η. This algorithm can be improved by adjusting this term C automatically according to the instantaneous output error. This gives rise to the complex-valued error-adaptive normalised nonlinear gradient descent (CEANNGD) algorithm. Applying the same method as in (8) to derive the second order derivative we obtain

$$\frac{\partial e^2(k)}{\partial w_n(k) \partial w_m(k)} = -\Phi''(net(k)) x_n(k) x_m(k). \qquad (11)$$

Taking the second and higher order terms to be the remainder of the expansion, $R_n(k)$ and recognising that $net(k) = \mathbf{x}^T(k)\mathbf{w}(k)$ we can compute $R_n(k)$ using the mean value theorem (MVT) and Rolle's theorem to give

$$R_n(k) = -\eta e(k) \hat{C}(k), \qquad (12)$$

such that

$$\begin{aligned} \hat{C}(k) = &\frac{1}{2!} \eta e(k) |\Phi'[\mathbf{x}^T(k)\mathbf{w}(k)]|^2 \\ &\times \Phi''[\mathbf{x}^T(k)(\mathbf{w}(k) + \theta \Delta \mathbf{w}(k))]||x_n(k)||_2^4, \end{aligned} \qquad (13)$$

and $0 < \theta \leq 1$. The expanded error term (9) can now be written as

$$\begin{aligned} e(k+1) = &[1 - \eta |\Phi'[\mathbf{x}^T(k)\mathbf{w}(k)]|_2^2 ||\mathbf{x}(k)||_2^2 \\ &- \frac{1}{2!} \eta^2 e(k) |\Phi'[\mathbf{x}^T(k)\mathbf{w}(k)]|^2 \\ &\times \Phi''[\mathbf{x}^T(k)(\mathbf{w}(k) + \theta \Delta \mathbf{w}(k))]||\mathbf{x}(k)||_2^4] e(k). \end{aligned} \qquad (14)$$

It is shown in [5] that explicit computation of $\hat{C}(k)$ has a large computational burden. Therefore it is proposed that we re-write the Taylor series expansion in terms of a function $\gamma(k)$ such that the squared expansion of the instantaneous output error becomes

$$e^2(k+1) = \gamma^2(k)e^2(k), \tag{15}$$

and $\gamma(k)$ is continuous on the set [5]

$$\bar{\mathbb{R}} = \mathbb{R} \setminus \{-|\Phi'(net(k))|^2||\mathbf{x}||_2^2\}, \tag{16}$$

where $\mathbb{R}$ is the set of real numbers. It is then shown that $\hat{C}$ is bounded as

$$\frac{-|\Phi'(net(k))|^2||\mathbf{x}(k)||_2^2}{2} \leq \hat{C}. \tag{17}$$

To ensure that a true convergence is obtained, we wish to derive some update term such that $C(k) = C(k-1) + \Delta C(k)$ for $k = 1, 2, \ldots$. However, for signals with rich dynamics and long time correlation, explicit computation of C can lead to large perturbation in the learning rate. To this cause, an iterative algorithm is proposed in [5] that employs the proportional squared error as the instantaneous gradient. In terms of a complex-valued algorithm this can be interpreted as

$$C(k) = \alpha C(k-1) + \mu\zeta^2(k) \tag{18}$$

where the function $\zeta(k) = |e(k)|$ provides the algorithm with the instantaneous gradient, the smoothing operator $\alpha = 1$, and $\mu \in \mathbb{R}$ is chosen to be a small positive constant which denotes the step size of the algorithm.

It can be shown that the CEANNGD algorithm can cause large fluctuations of the learning rate, $\eta(k)$, as the algorithm reaches the optimal point. At this point the error term tends to oscillate causing the learning rate to contain significant misadjustment errors. To this cause, following the proposed method given in [6], we update term for $C(k)$ in the CNNGD algorithm. This method incorporates a smoothing operator into the update term of the algorithm, in this case $\zeta(k)$ is defined as,

$$\zeta(k) = \kappa\zeta(k-1) + (1-\kappa)|e(k)e(k-1)|, \tag{19}$$

where $\kappa \in \mathbb{R}$ is a weighting parameter that controls the time averaging constant. It can be seen from (19) that consecutive terms of the instantaneous output error are multiplied, this is to minimize uncorrelated noise in the update of C. To this cause, it is proposed that the smoothed CEANNGD (SCEANNGD) algorithm be defined with $0 < \alpha < 1$ which controls the smoothing of the update term and μ is some small positive constant that controls the speed of convergence of $C(k)$. To this cause, the term $C(k)$ in the denominator of the learning rate is adjusted according to the filtering of the instantaneous output errors. This way, the analysis of the term $C(k)$ in the CEANNGD and SCEANNGD algorithms, conforms to the analysis given in [6].

5. EXPERIMENTAL RESULTS

To investigate the performance of the complex-valued error-adaptive normalised nonlinear gradient descent (CEANNGD) and smoothed CEANNGD (SCEANNGD) algorithms compared to the complex-valued normalised nonlinear gradient descent (CNNGD) and the complex-valued nonlinear gradient descent (CNGD) algorithms, they were applied to the problem of time-series prediction of complex-valued coloured and nonlinear signals. In all the experiments the order of the filter was $N = 10$ and the holomorphic activation function $\Phi(\cdot)$ was chosen to be the complex-valued hyperbolic tangent function

$$\Phi(z, \beta) = \frac{e^{\beta z} - e^{-\beta z}}{e^{\beta z} + e^{-\beta z}} \tag{20}$$

where $\beta = 0.2$ controls the slope of the algorithm and $z \in \mathbb{C}$. The first experiment involved averaging the performance curves of 700 independent simulations on the prediction of a complex coloured signal. The input, $r(k)$, was normally distributed $\mathcal{N}(0,1)$ complex white noise that was then passed through a stable AR filter defined as

$$\begin{aligned} y(k) &= 1.79y(k-1) - 1.85y(k-2) + 1.27y(k-3) \\ &\quad -0.41y(k-4) + r(k) \end{aligned} \tag{21}$$

In the CNGD algorithm a learning rate $\eta = 0.1$ was chosen, and in the CNNGD the values were chosen to be $C = 0.7$ and $C = 2$. In the SCEANNGD algorithm the values chosen were $\kappa = 0.5$, $\mu = 0.1$, $\alpha = 0.97$, and $\alpha = 0.3$. Figure

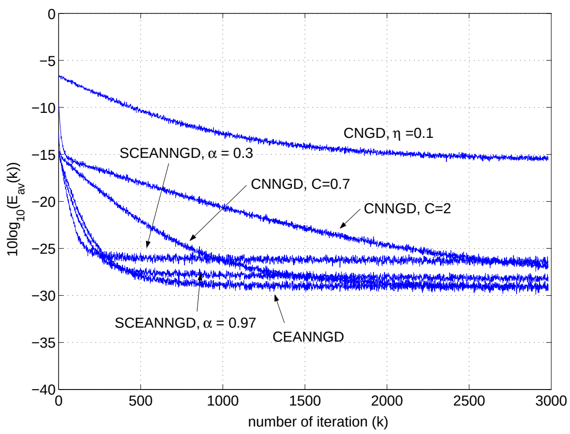

Fig. 2. Averaged convergence curves for CNGD, CNNGD, CEANNGD, and SCEANNGD on a coloured input signal

2 shows the performance curves for the CNGD, CNNGD,

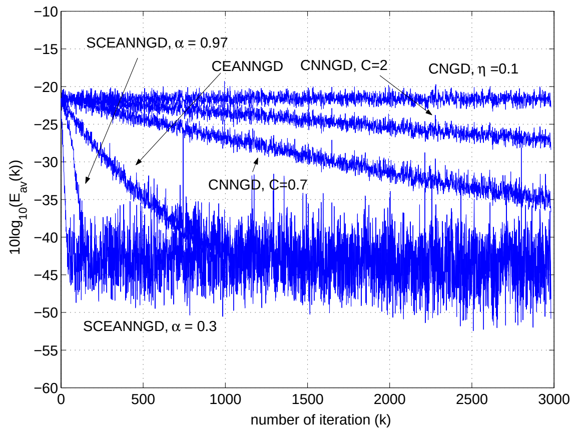

Fig. 3. Averaged convergence curves for CNGD, CNNGD, CEANNGD, and SCEANNGD on a nonlinear input signal

CEANNGD, and SCEANNGD algorithms on coloured input. A quantitative performance was based on E_{av} which is the average value of the squared modulus error. The CNGD algorithm reached $E_{av} = -15.0\ dB$ after $3,000$ iterations whereas the CNNGD algorithm with a value of $C = 2.0$ reached $E_{av} = -26.0\ dB$ after $3,000$ iterations of the algorithm. Similarly, the CNNGD algorithm with a value of $C = 0.7$ reached $E_{av} = -28.0\ dB$ after $2,500$ iterations. The CEANNGD algorithm converged to $E_{av} = -28.0\ dB$ after just $1,000$ iterations, which shows a significant increase in performance. The SCEANNGD algorithm converged to$E_{av} = -27.5\ dB$ in 1000 iterations and $E_{av} = -26.0\ dB$ in 1000 iterations for values of $\alpha = 0.97$ and $\alpha = 0.3$ respectively. The second experiment consisted of averaging the performance curves of 700 independent simulations on the prediction of a nonlinear time-series. The input, $r(k)$, was normally distributed $\mathcal{N}(0,1)$ complex white noise that was then passed through a benchmark nonlinear filter defined as

$$y(k) = \frac{y(k-1)}{1+y^2(k-1)} + r^3(k) \qquad (22)$$

In the CNGD algorithm a learning rate $\eta = 0.1$ was chosen, and in the CNNGD algorithms, values of $C = 2$ and $C = 0.7$ were chosen. In the SCEANNGD algorithm the values chosen were $\kappa = 0.5$, $\mu = 0.1$, $\alpha = 0.97$, and $\alpha = 0.3$. Figure 3 shows the performance curves for the CNGD, CNNGD, CEANNGD, and SCEANNGD algorithms. The CNGD algorithm reached $E_{av} = -21.0\ dB$ after $3,000$ iterations of the algorithm. The CNNGD algorithm for a value of $C = 2$ reached $E_{av} = -27.0\ dB$ after $3,000$ iterations, whereas if $C = 0.7$ the CNNGD reached $E_{av} = -35.0\ dB$ after $3,000$ iterations. The CEANNGD algorithm converged to a value of $E_{av} = -44.0\ dB$ after just $2,000$ iterations of the algorithm. The SCEANNGD algorithm converged to$E_{av} = -44.0\ dB$ in 400 iterations and $E_{av} = -44.0\ dB$ in 100 iterations for values of $\alpha = 0.97$ and $\alpha = 0.3$ respectively. In the experiment of nonlinear prediction, it is clear that there is some noise in the performance curves for all algorithms. In this way the SCEANNGD outperforms the CEANNGD by filtering the instantaneous output errors, which in turn provides a more accurate estimate of $C(k)$. However, for signals with uniform distribution and high signal to noise ratio, the CEANNGD algorithm outperforms the SCEANNGD algorithm.

6. CONCLUSIONS

A complex-valued error-adaptive normalised nonlinear gradient descent (CEANNGD) and smoothed CEANNGD (SCEANNGD) algorithm for complex-valued nonlinear FIR filters has been derived. It has been shown that the resulting algorithms are an extension of the fully adaptive normalised nonlinear gradient descent (FANNGD) algorithm given in [5] to the complex plane via the use of holomorphic activation functions. As such, they have been shown to be better suited for the processing of nonlinear and nonstationary signals. Experimental results have shown that the CEANNGD and SCEANNGD algorithms outperform the complex-valued nonlinear gradient descent (CNGD) algorithm and the complex-valued normalised nonlinear gradient descent (CNNGD) algorithm on the prediction of complex-valued coloured and nonlinear signals.

7. REFERENCES

[1] B. Widrow, J. McCool, and M. Ball, "The Complex LMS Algorithm," *Proceedings of the IEEE*, vol. 63, pp. 719–720, 1975.

[2] T. Kim and T. Adali, "Complex Backpropagation Neural Network Using Elementary Transcendental Activation Functions," *In Proceedings of IEEE ICASSP'01, Salt Lake City*, pp. 1281–1284, 2001.

[3] A. I. Hanna and D. P. Mandic, "A Normalised Complex Backpropagation Algorithm," *In Proceedings of IEEE ICASSP'02, Orlando*, pp. 977–980, 2002.

[4] D. P. Mandic and J. A. Chambers, *Recurrent Neural Networks for Prediction*, John Wiley and Sons, 2001.

[5] I. Krcmar and D. P. Mandic, "A Fully Adaptive Normalized Nonlinear Gradient Descent Algorithm for Nonlinear System Identification," *in Proceedings of ICASSP 2001*, vol. 6, pp. 3493–3496, 2001.

[6] T. Aboulnasr and K. Mayyas, "A Robust Variable Step-Size LMS-Type Algorithm: Analysis and Simulations," *IEEE Transactions on Signal Processing*, vol. 45, no. 3, pp. 631–639, 1997.