Application of Neural Adaptive Filters for Analysis and Modeling of Complex Valued Electricity Load Time Series

Slavisa Krunic
Siemens d.o.o. Banja Luka
Banja Luka, BiH
slavisa.krunic@siemens.com

Igor R. Krcmar, Petar S. Maric, Milorad M. Bozic
Faculty of Electrical Engineering
University of Banja Luka
Banja Luka, BiH
{ikrcmar, pmaric, mbozic}@etfbl.net

***Abstract*— Increased energy demand, together with deregulated energy market, emphasizes the role of accurate electricity load prediction. Design of the reliable predictor has to capture the nature of the electricity load time series. Further, electricity load time series requires modeling in complex domain. Design of neural networks based models for electricity load prediction tasks demands appropriate choice of the activation function of a neuron, structure and size of the training set, and learning algorithm. Neural adaptive filters, with their inherent simplicity and efficient learning algorithms, provide adequate tool for analysis and modeling of electricity load time series. Also, application of collaborative adaptive filters can give a deeper insight in the nature of electricity load time series, thus enabling better predictions. Experiments, carried out on the test load signal, metered on a medium voltage feeder. support the analysis.**

Keywords- collaborative adaptive filters, gradient descent neural adaptive filters, electricity load modeling, electricity load prediction

I. Introduction

The role of accurate electricity load time series prediction is emphasized by increase of energy demand and deregulated energy market. In oreder to obtain an accurate prediction, efficient and reliable predictor has to be employed. Further, the predictor must reflect the nature of time series at hand. Neural network (NN) models has ability to cope with process nonlinearity and to learn long term dependencies [1]. Also, data preprocesing and adequate learning algorithm can help when NN has to deal with nonstationary time series. However design of NN based predictor requires answers to several questions. The questions regard choice of appropriate activation function (AF), learning algorithm, size and structure of training set, and NN structure [1]. A class of neural adaptive filters, equipped with gradient descent (GD) based learning algorithm, due to their simple structure, can not provide accurate long term predictions, yet they might provide insight in the caracter of the time series [2]. Also, they can indicate answers to some of the above mentioned questions, i.e. which AF of a neuron to apply, what is appropriate learning algorithm, as well as, to indicate the size of training set. Physical nature, of a power consumption process, requires modeling in the complex domain. The main challenge, when developing NN based model for prediction of complex valued time series, lies in a search for an appropriate nonlinear AF. Due to the result of the Liouville's theorem, bounded and analytic complex nonlinear function on the whole complex plane C does not exist [3-5]. This fact led to several approaches. Within the dual univariate AF (DUAF) setup [3], real and imaginary parts of complex valued time series are processed separately. So, there are two real-valued bounded and analytic nonlinear AFs. Thus, weight update is split in two part. The split complex (SC) approach means that real and imaginary part of the complex net input are processed separately by real valued sigmoid functions [3]. So, there is one weight update equation within the SC approach. Application of meromorphic functions as AFs defines the fully complex (FC) approach. Meromorphic functions are analytic everywhere, except on a discrete subset of C, which consists of function singularities [3,4]. Our aim is to investigate are the neural adaptive filters suitable for analysis and modeling of complex valued electricity load time series. To this cause, we provide comparative analysis of different GD based learning algorithms, as well as, different structures of neural adaptive filters. The analysis is carried out on the metered values of a complex-valued energy, in a distribution, medium voltage, grid.

II. Neural adaptive filters

The structure of a neural adaptive finite impulse response (FIR) filter is given on Fig. 1. Operation of the filter (Fig. 1) is described as follows [6,2,3]

$$y(k) = \Phi(net(k)) \quad (1)$$

$$net(k) = \mathbf{x}^T(k)\mathbf{w}(k) \quad (2)$$

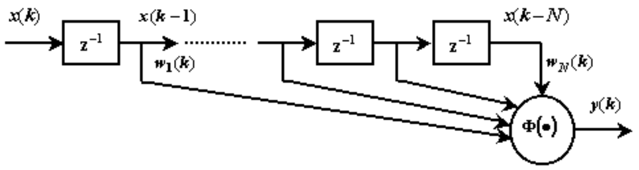

Figure 1. A neural adaptive FIR filter.

where *y(k)* denotes filter output, Φ(•) denotes AF of the output neuron, *net(k)* is input to the neuron, *N* is length of filter tap inputs, *k* is a discrete time instant, $\mathbf{w}(k)=[w_1(k), w_2(k),\ldots, w_N(k)]^T$ is the filter weight vector, $(\bullet)^T$ denotes vector transpose and $\mathbf{x}(k)=[x_1(k), x_2(k),\ldots, x_N(k)]^T$ is the filter input vector, where $x_i(k)= x(k-i),\ i=1,2,\ldots,N$. Gradient descent learning algorithm for the filter (Fig. 1), is described by

$$\mathbf{w}(k+1)=\mathbf{w}(k)-\mu\nabla_{\mathbf{w}}J(k) \quad (3)$$

$$J(k)=\frac{1}{2}|e(k)|^2 \quad (4)$$

$$e(k)=d(k)-y(k) \quad (5)$$

where μ is the step size, *J(k)* is the cost function, $\nabla_{\mathbf{w}}(\bullet)$ denotes gradient of a scalar function with respect to the weight vector **w**, $|\bullet|$ denotes absolute value, *e(k)* is the error at the output neuron and *d(k)* is some desired, teaching signal. Computation of the gradient $\nabla_{\mathbf{w}} J(k)$ depends on the type of nonlinear AF. According to [3,6] we can distinguish between following cases.

A. FC FIR adaptive filter

Computation of the cost function (4) gradient, having in mind that Φ is a meromorphic function, is as follows

$$\nabla_{\mathbf{w}}J(k)=\nabla_{\mathbf{w}_r}J(k)+j\nabla_{\mathbf{w}_i}J(k). \quad (6)$$

where $j=\sqrt{-1}$ and $\mathbf{w}_r=\Re(\mathbf{w})$, $\mathbf{w}_i=\Im(\mathbf{w})$. If we introduce $e(k)=\Re(e(k))+j\Im(e(k))=e_r(k)+je_i(k)$ we have

$$\begin{aligned}\nabla_{\mathbf{w}}J(k)&=-\mathbf{x}^*(k)\left[\Phi'^*(k)e_r(k)+j\Phi'^*(k)e_i(k)\right]\\&=-e(k)\mathbf{x}^*(k)\Phi'^*(k).\end{aligned} \quad (7)$$

where for convenience $\Phi'^*(net(k))=\Phi'^*(k)$. Further, $(\bullet)'$ denotes first derivative and $(\bullet)^*$ denotes complex conjugate. Now, from (3) and (7) we have weight update equation for the complex nonlinear gradient descent (CNGD) algorithm

$$\mathbf{w}(k+1)=\mathbf{w}(k)+\mu e(k)\Phi'^*(k)\mathbf{x}^*(k). \quad (8)$$

If Φ is linear AF, i.e. Φ=*net*(*k*), (8) becomes

$$\mathbf{w}(k+1)=\mathbf{w}(k)+\mu e(k)\Phi'^*(k)\mathbf{x}^*(k). \quad (9)$$

which defines weight update for the complex least mean squares (CLMS) algorithm.

B. SC FIR adaptive filter

Within real-imaginary SC approach, output of the filter (1) is defined as [8]

$$\begin{aligned}y(k)&=\Phi(\mathbf{x}^T(k)\mathbf{w}(k))=\Phi(net(k))\\&=\sigma(net_r(k))+j\sigma(net_i(k)).\end{aligned} \quad (10)$$

where $net(k)=\Re(net(k))+j\Im(net(k))=net_r(k)+jnet_i(k)$ and σ denotes real-valued sigmoid nonlinearity. Computation of the gradient (6), for the weight update (3) gives

$$\begin{aligned}\mathbf{w}(k+1)=\mathbf{w}(k)&+\mu[e_r(k)\sigma'(net_r(k))\\&+je_i(k)\sigma'(net_i(k))]\mathbf{x}^*(k).\end{aligned} \quad (11)$$

Note that if σ is linear function (11) transforms into (9)..

C. DUAF approach

Within the DUAF approach signal processing is performed parallel on two real-valued signals. In this case signal is processed by two LMS, in linear case, or NGD algorithms, in nonlinear case.

D. Normalized learning algorithms

Normalized gradient learning algorithms provide variable step size μ., which yields Within the normalized complex least mean squares (NCLMS) algorithm, the step size is $\mu_{NCLMS}(k)=\eta/(\varepsilon+\|\mathbf{x}(k)\|_2^2)$, while in a nonlinear case, normalized complex nonlinear gradient descent (NCNGD) algorithm has the step size $\mu_{NCNGD}(k)=\eta/(C+|\Phi'(k)|^2\|\mathbf{x}(k)\|_2^2)$. In both cases η, ε, and C are small positive constants.

III. COLLABORATIVE ADAPTIVE FILTERS

Structure of the collaborative adaptive filter [3] is given on the Fig. 2. The idea is to form a convex combination of two FIR adaptive filters. Even it is, at the first, intended for improvement of the performance of standard algorithms, it can be used to compare different algorithms and structures.

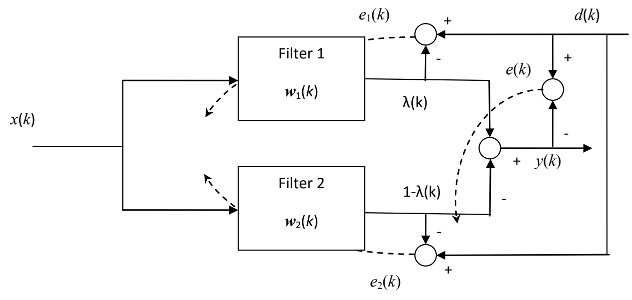

Figure 2. Structure of the collaborative adaptive filters.

Operation of the structure given on the Fig. 2. can be described as follows

$$y(k)=\lambda(k)y_1(k)+(1-\lambda(k))y_2(k) \quad (12)$$

where $0 \le \lambda(k) \le 1$, $y_1(k)$ and $y_2(k)$ denote outputs of filter 1 and filter 2, respectively. Parameter $0 \le \lambda(k) \le 1$ is updated on-line, in the GD manner, thus we have

$$\begin{aligned} \lambda(k+1) &= \lambda(k) - \mu_\lambda \nabla_\lambda J(k) \\ &= \lambda(k) + \mu_\lambda e(k)(y_1(k) - y_2(k)) \end{aligned} \tag{13}$$

In the case of complex valued signals (13) becomes

$$\begin{aligned} \lambda(k+1) &= \lambda(k) - \mu_\lambda \nabla_\lambda J(k) \\ &= \lambda(k) + \mu_\lambda e(k)(y_1(k) - y_2(k))^* \end{aligned} \tag{14}$$

From (13) and (14) follows application of collaborative adaptive filters in algorithm analysis. Change of the value of $\lambda(k)$ indicates which algorithm has better performance, and therefore more suitable for concrete application.

IV. Experimental Results

The experiments were carried out, as one step ahead signal prediction, in order to analyze application different learning algorithms and structures in complex valued electricity load time series. The normalized test complex valued load signal is shown on the Fig. 4. and Fig. 5. The signals represent fifteen minutes average of active and reactive power, metered at the 10 kV feeder, in the Transformer station Banja Luka 2. The signals were normalized with respect to the maximum of the load absolute value, in order to fit the range of sigmoid nonlinearity. The logistic AF $\Phi(z)=1/(1+\exp(-\beta z))$ was used within the experiments as sigmoid nonlinearity at the neuron, with the slope β=4. The performance measure was standard prediction gain $R_p=10\log_{10}(\sigma_y^2/\sigma_e^2)$, where σ_y^2 and σ_e^2 denote variances of the predicted signal and the output error, respectively. In the first experiment the step size was μ=0.3. The length of filter tap inputs varied from 1 to 30. In the NCLMS and NCNGD algorithms the parameters were set as η=0.3, ε=C=10^{-5}.. Summary of the results for a linear adaptive filters is shown on the Fig. 5, while results for a nonlinear neural adaptive filter are shown on the Fig. 6.

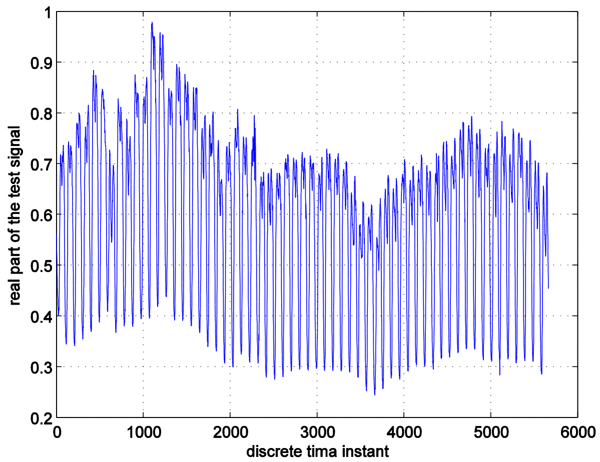

Figure 3. Real part of the test signal.

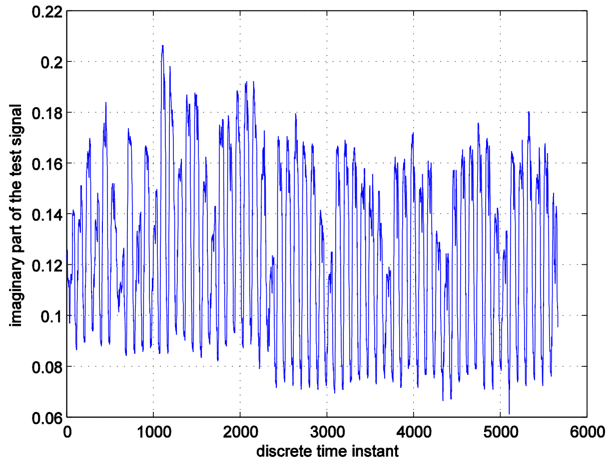

Figure 4. Imaginary part of the test signal.

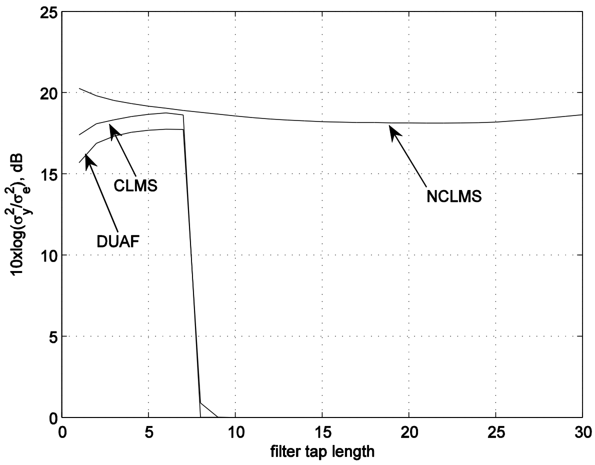

Figure 5. Performance of linear adaptive filters in the first experiment.

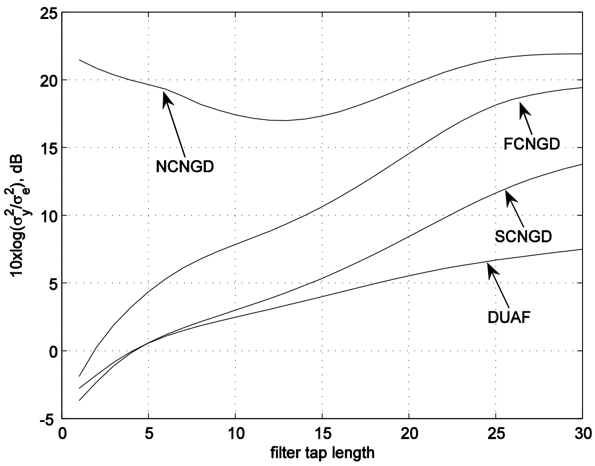

Figure 6. Performance of nonlinear adaptive FIR filters in the second experiment.

In the second experiment collaborative adaptive filters were applied for the complex valued load prediction task.

Three combinations of algorithms were tested, i.e. CLMS and CNGD, CLMS and NCLMS, and NCLMS and NCNGD. For all filters FC approach was applied. Filter tap lengths were *N*=10 and *N*=30. For higher filter tap lengths the step size of the CLMS was reduced from $\mu = 0.3$ to $\mu = 0.05$ due to stability problems. The results of the second experiment are summarized in the Table. I From the first experiment it is obvious that FC approach gives the best results. Further, increase of the filter order significantly improves performance of nonlinear algorithms. Also, CLMS algorithm and linear DUAF approach was become unstable for larger filter tap lengths. Normalized algorithms, both for linear and nonlinear algorithms, have the best performance. Results of the second experiment verify the outcome of the first experiment. They indicate presence of long term dependencies in the analyzed time series and its nonlinear nature. Even in the case where linear algorithm outperforms nonlinear one, overall performance of the collaborative structure is improved, comparing to the performance of its constituents. When the performance of nonlinear and/or normalized algorithms is at its maximum, companion algorithm decreases performance of the overall collaborative structure.

TABLE I. SUMMARY OF THE RESULTS IN THE SECOND EXPERIMENT

Filter Tap Length, N	**CLMS and FCNGD**			**CLMS and NCLMS**			**NCLMS and NCNGD**		
	CLMS	**FCNGD**	**Col. Filt.**	**CLMS**	**NCLMs**	**Col. Filt.**	**NCLMS**	**NCNGD**	**Col. Filt.**
10	**15.30dB**	**9.30dB**	**17.73dB**	**15.30dB**	**18.17dB**	**17.45dB**	**18.17dB**	**17.64dB**	**21.92dB**
	η=0.1	**η=0.3**	**η_λ=0.5**	**η=0.1**	**η=0.3**	**η_λ=0.5**	**η=0.3**	**η=0.3**	**η_λ=0.5**
30	**15.13dB**	**19.44dB**	**19.11dB**	**15.13dB**	**18.47dB**	**17.73dB**	**18.47dB**	**21.92dB**	**21.83dB**
	η=0.05	**η=0.3**	**η_λ=0.5**	**η=0.05**	**η=0.3**	**η_λ=0.5**	**η=0.3**	**η=0.3**	**η_λ=0.5**

V. CONCLUSIONS

Complex valued electricity load time series prediction is very important for operation of power utilities in deregulated energy market. Nature of the electricity load time series demands modeling in the complex domain. Development, of the NN based model for the load prediction tasks, requires choices to be made on the appropriate AF of a neuron, learning algorithm, and size and structure of the training set. To this cause, GD neural adaptive filters have been employed in an analysis and modeling of electricity load time series. Filters with larger tap length have had a better performance thus indicating long term dependencies of analyzed time series. Further, performance of nonlinear and/or normalized algorithms indicates nonlinear nature of analyzed time series. Also, filters with FC AF have exhibited increased performance, thus indicating that the AF of a neuron should be a meromorphic function.

REFERENCES

[1] S. Haykin, *Neural networks - A comprehensive foundation*. Prentice Hall, 1994.

[2] D. P. Mandic and J. A. Chambers, *Recurrent Neural Networks for Prediction: Learning Algorithms, Architectures and Stability*, John Wiley & Sons, 2001.

[3] D. P. Mandic and S. l. Goh, *Complex valued nonlinear adaptive filters: noncircularity, widely linear and neural models*, John Wiley & Sons, 2009.

[4] T. Needham, *Visual Complex Analysis*, Oxford University Press, 1997.

[5] T. Kim and T. Adali, “Approximation by fully complex multilayer perceptrons,” *Neural Computations*, vol. 15(7), pp. 1641-1666, 2003

[6] Igor R. Krcmar, Petar S. Maric, and Milorad M. Bozic, “A Class of Neural Adaptive FIR Filters for Complex-Valued Load Prediction,” Proceedings of 10th Seminar on Neural Network Applications in Electrical Engineering (NEUREL-2010), pp. 37–39, 2010.