ISSN : 0976-8491 (Online) | ISSN : 2229-4333 (Print)

An amplified Echo State Network for Nonlinear Adaptive Filtering of intricate Noncircular Signals

V. V. Durga Bhavani
CSE from Jawharlal Nehru University

Abstract
This combination of augmented intricate statistics and enhanced adaptively within ESNs also facilitates the processing of bivariate signals with strong component correlations. A narrative intricate Echo State Network (ESN), utilizing full second-order statistical information in the intricate domain, is introduced. This is achieved through the use of the so-called augmented intricate statistics, thus making intricate ESNs suitable for processing the generality of complex-valued signals, both second-order circular (proper) and noncircular (improper). Next, in order to deal with nonstationary processes with large nonlinear dynamics, a nonlinear readout layer is introduced and is further equipped with an adaptive amplitude of the nonlinearity. Simulations in the prediction setting on both circular and noncircular synthetic benchmark processes and real-world noncircular and nonstationary wind signals support the analysis.

Keywords
Augmented Intricate Statistics, Intricate Non-Circularity, Echo State Networks, Widely Linear Modeling, Wind Prediction

I. Introduction

RECURRENT neural networks (RNNs) are a class of nonlinear adaptive filters with feedback, whose compu-tational power stems from their ability to act as universal approximators for any continuous function on a compact domain . Owing to their rich inherent memory through feedback, RNNs have found applications in the modeling of highly nonlinear dynamic systems and the associated attractor dynamics. They are typically used in the system identification time-series prediction and adaptive noise cancellation settings where for the nonstationary and nonlinear nature of the signals and typically long impulse responses, using the class of static feedforward networks or transversal filters would result in undermodeling Recently, a class of discrete-time RNNs, called Echo State Networks (ESNs), have been introduced, with the aim to reduce the complexity of computation encountered by standard RNNs. The principle behind ESNs is to separate the RNN architecture into two constituent components: a recur-rent architecture, called the “dynamical reservoir” or “hidden layer,” and a memoryless output layer, called the “readout neuron.”

The recurrent architecture consists of a randomly generated group of hidden neurons with a specified degree of recurrent connections, and should satisfy the so-called “echo state property” to maintain stability . This way, the high computational complexity of RNNs is significantly reduced due to the sparse connections among the hidden neurons, in addition, the learning requirements are reduced to only the weights connecting the hidden layer and the readout neuron. Many real-world bivariate processes, such as vector fields and directional signals with “intensity” and “direction” com-ponents, are most conveniently represented when considered complex-valued . Consequently, in the neural network literature, several important approaches have been extended to the complex domain, examples include coherent neural net-works for sensorimotor systems , sonar signal prediction and image enhancement by multivalued neurons , gray-scale image processing by complex-valued multistate neural associate memory , and geometric figure transformation via complex-valued backpropagation networks.

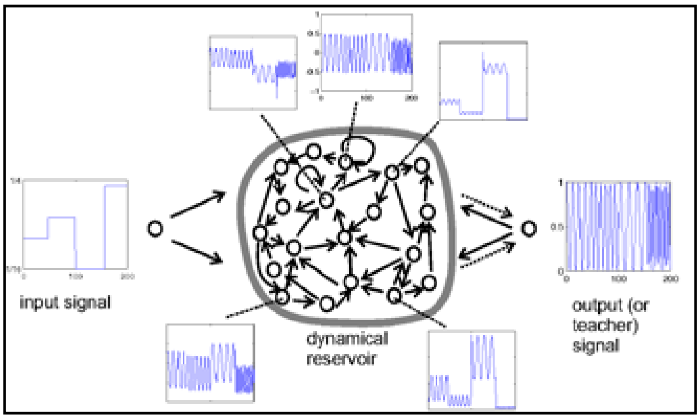

Fig. 1: The Basic Schema of an ESN, Illustrated with a Tuneable Frequency Generator Task. Solid Arrows Indicate Fixed, Random Connections; Dotted Arrows Trainable Connections

The first extension of ESNs into the complex domain C was proposed in , this network had a linear output mapping, and was trained by the complex-valued Wiener filter, thus making the network second-order optimal for the processing of circular stationary data. Results in adaptive filtering dealing with real-world complex-valued data suggest that, due to the linearity of the output mapping, the degree of universal func-tion approximation exhibited by standard ESNs may not be sufficient. To that end, a nonlinear output layer within ESNs, i.e., the linear mapping followed by a nonlinear activation function, has been proposed in. To deal with common problems experienced in neural network training, such as saturation and slow convergence resulting from the unknown and large dynamics of inputs, the nonlinear output layer of ESNs has further been equipped with an adaptive amplitude of the nonlinearity.

Adaptive filtering algorithms in the complex domain C are usually considered generic extensions of their real domain counterparts. For instance, a common assumption explicitly or implicitly exists in the signal processing literature that complete second-order statistical information of a zero-mean complex vector z is contained in the covariance matrix. However, recent results in so-called augmented complex statistics show that, in general, this leads to suboptimal estimation and that for the fenerality of complex valued random process both the covariance T matrix and the pseudo-covariance matrix should be considered to completely capture the second order statistics available. In practicem this is achieved by widely linear modeling, which has been proved to be particularly advantageous when processing secondorder noncircular(improper) signals for which the probability distributions are not rotation invariant.

Recently, augmented complex statistics have been intro-duced into several key learning algorithms, examples include the augmented complex least means square (ACLMS) [23], augmented complex extended Kalman filter , and aug-mented complex real-time recurrent learning. Following on these results, we here introduce augmented statistics into the training of complex ESNs, allowing us to make use of all the available second-order statistical information, and to produce optimal estimates for second-order noncircular (improper) signals.

This paper is organized as follows. In Section II, we provide an overview of widely linear estimation and second-order augmented complex statistics. In Section III, the augmented complex ESN and its nonlinear variants are derived. Simulations on both synthetic circular and noncircular signals and real-world nonstationary and noncircular wind signals are given in Section IV, demonstrating the advantage of the augmented ESN over standard complex ESNs. Finally, Section V, concludes this paper.

II. Widely Linear Modeling

Consider the real-valued Mean Squared Error (MSE) estimator

yö = E[y|x] (1)

which estimates the values of signal y in terms of another observation x. For zero-mean jointly normal y and x, the linear model solution is

yö = x^T h (2)

where h = [h 1, ... , h N $]^T$ is a vector of fixed filter coefficients, and the past of the observed variable is contained in the regressor vector x = [x_1, ... , x_N $]^T$.

In the complex domain, it is assumed that we can use the same form of conditional mean estimator that for real-valued signals given in (1), leading to the standard complex linear Minimum Mean Squared Error (MMSE) estimator

yö = z^H h (3)

where the symbol (á)H denotes the Hermitian transform oper-ator. However, the real-valued linear estimator in applies to both the real and imaginary parts of complex variables

$$y\ddot{o}_r = E[y_r | z_r , z_i]$$
$$y\ddot{o}_i = E[y_i | z_r , z_i]. \quad (4)$$

A more general MSE estimator than that in (3) can be expressed as

$$y\ddot{o} = E[y_r | z_r , z_i] + ♫ E[y_i | z_r , z_i]. \quad (5)$$

leading to a widely linear estimator for complex-valued data, given by

y = z^T h + z^H g (6)

where h and g are complex-valued coefficient vectors. This estimator is suitable for linear MMSE estimation of the generality of complex-valued processes (both circular and noncircular), as it accounts for complete second-order information in C, as shown below.

A. Second-Order Augmented Complex Statistics

It is clear that the covariance matrix C_{zz} = E[zz^H] alone does not have sufficient degrees of freedom to describe full second-order statistics in C [20] and, in order to make use of all the available statistical information, we also need to con-sider the pseudo-covariance matrix P_{zz} = E[zz^T]. Processes whose second-order statistics can be accurately described by only the covariance matrix, i.e., those for which the pseudo-covariance P_{zz} = 0, are termed second-order circular (or proper). In general, the notion of circularity extends beyond second-order statistics to describe the class of signals with rotation-invariant distributions P[á] for which P[z] = P[$ze^{j\theta}$] for θ ∈ [0, 2π). In most real-world applications, complex signals are second-order noncircular or improper, and their probability density functions are not rotation-invariant. In practice, to account for the improperness, the input vector z is concatenated with its conjugate z□, to produce an augmented 2N × 1 input vector.

III. Augmented ESN

A. Standard Complex ESN with a Linear Output Mapping

Fig. 1, shows the architecture of a standard ESN, which is composed of K external input neurons, L readout neurons, and N internal units. Without loss of generality, we shall address ESNs with one readout neuron (L = 1), as this facilitates the nonlinear adaptive filtering setting within the ESN architecture. The input and internal weights are stored, respectively, in the (N ×K) and (N ×N) weight matrices W_{ip}, W_{in}, vector w b comprises the feedback weights connecting the readout neuron and the internal units, vector x(k) is the (N ×1) internal state vector, u(k) represents the (K ×1) input vector, and y(k) is the overall output. The network state at time instant k, denoted by q(k), is a concatenation of the input u(k), internal state x(k), and the delayed output y(k − 1)

q(k)= ✢u(k), ... , u(k−K+1), x_1(k),..., x_N(k), y(k−1) T (10)

where as the internal unit dynamics are described by [12] x(k) = f'W_{ip} u(k) + W_{in} x(k − 1) + w_b y(k − 1) (11)

where f(á), is a vector-valued nonlinear activation function of the neurons within the reservoir.

The echo state property is provided by randomly choosing an internal weight matrix W_{in} and performing scaling to make the spectral radius ρ(W_{in}) < 1, thus ensuring that the network is stable, the input and feedback weights can be initialized arbitrarily [12]. For an ESN with a linear output mapping, the output y(k) is given by

y(k)=q^T(k)w(k) (12)

where w(k), is the weight vector corresponding to the output layer. Its update can be performed, e.g., based on the CLMS algorithm,

B. Augmented Complex ESN

With a Linear Readout Neuron based on the widely linear model given in Section II, we shall now derive the augmented widely linear stochastic gradient algorithm for the training of complex ESNs, thus making them suitable for processing general complex-valued signals(both circular and noncircular).

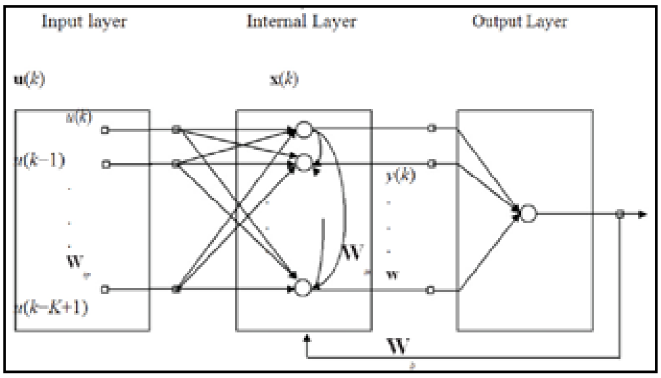

Fig. 2:

Due to the specific properties of the ESN architecture, the output y(k) of the augmented ESN with linear output mapping is governed by an asymmetric version of the widely linear model Note that the ESN has a local feedback (from the output to the internal state) and thus, unlike standard feedback structures the output within the state vector of the ESN does not require augmentation with its conjugate. Therefore, due to local feedback, the conjugate weight vector g(k) is only associated with the conjugate input signal. The standard and conjugate weight vectors are thus of different dimensions, but as with all widely linear models, the conjugate weight vector g(k) = 0 for a circular input signal.

C. ESNs with an Adaptive Amplitude of Output Nonlinearity

The nonlinear output layer has been introduced into ESNs to prove a sufficient degree of nonlinearity for enhanced modeling; however, this doesn't automatically guarantee optimal modeling, as some parameters, such as the amplitude of the nonlinear readout neuron, need to be chosen empirically. To this end, we now introduce gradient adaptive amplitude into the output nonlinearity of ESNs.

D. Qualities of Nonlinearity and Widely Linear Model

To illustrate the effect of nonlinearity and the widely linear model, we generated a circular doubly white noise n(k), with zero mean and unit variance, and passed it through a complex valued tanh nonlinearity, and the widely linear model given by
x(k) = WL(n(k)) = 0.6n(k) + 0.8n(k).
shows that the application of tanh nonlinearity alters the character of distribution, which cannot be achieved by using the widely linear model. It is therefore natural to combine the widely linear model and nonlinear processing in order to deal simultaneously with various aspects of the nature of the data. By introducing the adaptive amplitude of nonlinearity, we have an additional degree of freedom, allowing the nonlinear function to operate in a quasi-linear range if so required by the nature of the data.

IV. Simulations

To verify the potential of the proposed augmented ESNs compared to standard complex ESNs, we performed simulations on both benchmark synthetic proper and improper signals, and for noncircular real-world wind data.

For all signals, experiments were undertaken by averaging 200 independent simulation trials in the adaptive prediction setting. The nonlinearity at the nonlinear output layer of the ESNs was chosen to be the complex tanh function

$$(x) \qquad \frac{e^{\beta x} - e^{-\beta x}}{+ \quad -} \quad (45)$$

with slope β = 1. Ten neurons were used in the hidden layer, with the internal connection weights having 5% degree of connectivity. The input tap length was K = 1, with no bias input. The values of the randomly selected input as well as internal and feedback weights Wip , Win , and wb were taken from a uniform distribution in the range [−1, +1], and the spectral radius ρ(Win) was set to be 0.8. The learning rate was μ = 0.005 for all the learning algorithms considered, with the initial amplitude for the AACNGD algorithm λ(0) = 1 and the step size of the adaptive amplitude update η = 0.2.

The benchmark circular signal was a stable linear auto regressive driven by complex-valued doubly circular white Gaussian noise n(k) with zero mean and unit variance.

The benchmark noncircular signal was a complex AR moving-average (ARMA) process, whose transfer function was a combination of the MA model in and the stable AR model.

Table 1: Percentage of Enhanced Performance of Augmented ESN Algorithms for Complex Noncircular Signals

	Noncircular ARMA	Ikeda map	Wind (low)	Wind (medium)	Wind (high)
Augmented ESN (ACLMS)	95.5%	98.5%	91.5%	94%	91%
Augmented ESN (Augmented CNGD)	95%	98%	90%	**95%**	94.5%
Augmented ESN (Augmented AACNGD)	93.5%	99%	89.5%	92.5%	92%

Parts of complex valued data as two independent channels, had the worst performance. For the circular AR(4) signal, the performance, the performance of the augmented complex ESN was similar to that of standard ESN. For the noncircular signals, there was a significant improvement in the prediction gain when the augmented ESN was employed. As desired, the advantage of the nonlinear output layer over the linear output mapping wasmore pronounced in the prediction of the nonlinear synthetic signal and nonlinear and nonstationary real world wind reservoir within an ESN, the augmented ESN cannot guarantee enhanced performance over its standard version in evey trial, however as illustrated in above table, on average in more than 90% of the trails the widely linear algorithms outperformed the corresponding standard ones.

To further illustrate the advantage of using augmented complex statistics within complex-valued ESNs, we compared the MSEs of both the augmented and standard ESNs with adaptive amplitude of nonlinearity for the prediction of the complex-valued synthetic nonlinear and noncircular Ikeda map and the noncircular wind (high) signal. In both cases, the augmented ESN with a nonlinear readout neuron trained by the augmented AACNGD outperformed its standard version.

We next investigated the influences of two parameters re-lated to the generation of the internal layer, the degree of con-nectivity, and the spectral radius ρ(Win) on the performance of standard and augmented ESNs. In all the cases, for the prediction of real-world wind (medium) and (high) signals the augmented ESN trained by the augmented AACNGD algorithm achieved the best performance and that for both learning strategies it is desirable to keep a low degree of connectivity within the reservoir. This conforms to the ESN theory [12] that a small degree of connectivity can perform a relative decoupling of subnetworks with rich reservoir dynamics.

The size of the dynamical reservoir is another important parameter that influences the performance of ESNs, as it reflects their universal approximation ability. An ESN with a larger reservoir size can learn the signal dynamics with a higher accuracy [40], on one-step-ahead prediction of the noncircular ARMA process in (47). This, however, applies to stationary signals, whereas for fast-changing nonstationary processes, the larger reservoir caused saturation of internal neurons, resulting in performance degra-dation, as for the prediction of the nonsta-tionary wind (medium) signal. Observe that, in all the cases, the augmented ESNs outperformed their standard counterparts.

In the final set of simulations, we considered multistepahead prediction of the noncircular and nonstationary wind (medium) and (high) data.

V. Conclusion

An augmented complex ESN has been introduced for nonlinear adaptive filtering of the generality of complex-valued signals. The proposed ESN has been derived based on augmented complex statistics, thus making it suitable for both second-order circular and noncircular signals. For generalities, a nonlinear output layer has been introduced, and to deal with signals with large dynamics, an adaptive amplitude has been introduced into the output layer of the augmented ESN. The proposed augmented ESNs have been shown to exhibit theoretical and practical advantages over their conventional counterparts. This has been verified through comprehensive simulations on both synthetic noncircular data and real-world wind measurements, and over a range of parameters.

References

[1] Eric Antonelo, the site administrator,"Reservoir Computing: Shaping Dynamics into Information", 2011, [Online] Available: http://www.reservoir-computing.org, visited on October 12, 2011.

[2] H. Jaeger,"Tutorial on training recurrent neural networks, covering BPPT, RTRL, EKF and the "echo state network" approach," Fraunhofer Institute AIS, St. Augustin-Germany, Tech. Rep., 2002.

[3] H. Jaeger.,"The "echo state" approach to analysing and training recurrent neural networks," GMD-German National Research Institute for Computer Science, Tech. Rep. 148, 2001.

[4] W. Maass, T. Natschlaeger, and H. Markram, "Real-time computing without stable states: A new framework for neural computation based on perturbations," Neural Computation, Vol. 14, No. 11, pp. 2531–2560, 2002.

[5] S. Scherer, M. Oubbati, F. Schwenker, G. Palm,"Realtime emotion recognition from speech using echo state networks", Artificial Neural Networks in Pattern Recognition, Vol. 5064, pp. 205– 216, 2008.

[6] D. P. Mandic, J. A. Chambers, Recurrent Neural Networks for Prediction: Learning Algorithms, Architectures and Stability. New York: Wiley, 2001.

[7] K. S. Narendra, K. Parthasarathy,"Identification and control of dynamical systems using neural networks", IEEE Trans. Neural Netw., Vol. 1, No. 1, pp. 4–27, Mar. 1990.

[8] C. H. Lee, C. C. Teng,"Identification and control of dynamic systems using recurrent fuzzy neural networks", IEEE Trans. Fuzzy Syst., Vol. 8, No. 4, pp. 349–366, Aug. 2000.

[9] S. Haykin, L. Li,"Nonlinear adaptive prediction of nonstationary signals", IEEE Trans. Signal Process., Vol. 43, No. 2, pp. 526–535, Feb. 1995.

[10] D. P. Mandic, J. A. Chambers, "Toward an optimal PRNN-based nonlinear predictor", IEEE Trans. Neural Netw., Vol. 10, No. 6, pp. 1435– 1442, Nov. 1999.

[11] S. A. Billings, C. F. Fung,"Recurrent radial basis function networks for adaptive noise cancellation", Neural Netw., Vol. 8, No. 2, pp. 273– 290, 1995.

[12] C.-M. Lin, L.-Y. Chen, D. S. Yeung,"Adaptive filter design using recurrent cerebellar model articulation controller", IEEE Trans. Neural Netw., Vol. 21, No. 7, pp. 1149–1157, Jul. 2010.

[13] J. L. Elman,"Finding structure in time", Cognitive Sci., Vol. 14, No. 2, pp. 179–211, Mar. 1990.

[14] W. Liu, I. Park, J. C. Príncipe,"An information theoretic approach of designing sparse kernel adpative filters", IEEE Trans. Neural Netw., Vol. 20, No. 12, pp. 1950–1961, Dec. 2009.

[15] H. Jaeger, H. Haas,"Harnessing nonlinearity: Predicting chaotic systems and saving energy in wireless communication", Science, Vol. 304, No. 5667, pp. 78–80, Apr. 2004.

[16] H. Jaeger,"The echo state approach to analyzing and training neural networks", German Nat. Res. Inst. Inform. Technol., Sankt Augustin, Germany, Rep. 148, 2002.

[17] D. P. Mandic, S. L. Goh,"Complex Valued Nonlinear Adaptive Filters", Noncircularity, Widely Linear and Neural Models. New York: Wiley, 2009.

[18] A. Hirose,"Complex-Valued Neural Networks", New York: Springer-Verlag, 2006.

[19] I. N. Aizenberg, N. N. Aizenberg, J. P. L. Vandewalle, "Multi-Valued and Universal Binary Neurons: Theory, Learning and Applications. Norwell, MA: Kluwer Academic, 2000.

[20] S. Jankowski, A. Lozowski, J. M. Zurada,"Complex-valued multi-state neural associative memory," IEEE Trans. Neural Netw., Vol. 7, No. 6, pp. 1491–1496, Nov. 1996.

[21] T. Nitta,"An extension of the back-propagation algorithm to complex numbers", Neural Netw., Vol. 10, No. 8, pp. 1391–1415, Nov. 1997.

[22] S. Seth, M. C. Ozturk, J. C. Príncipe,"Signal processing with echo state networks in the complex domain", in Proc. IEEE Workshop Mach. Learn. Signal Process., Thessaloniki, Greece, Aug. 2007, pp. 408–412.

[23] Y. Xia, D. P. Mandic, M. M. Van Hulle, J. C. Príncipe, "A complex echo state network for nonlinear adaptive filtering", in Proc. IEEE Workshop Mach. Learn. Signal Process., Cancun, Mexico, Oct. 2008, 404–408.

[24] F. D. Neeser, J. L. Massey,"Proper complex random processes with applications to information theory", IEEE Trans. Inform. Theory, Vol. 39, No. 4, pp. 1293–1302, Jul. 1993.

[25] B. Picinbono, P. Chevalier,"Widely linear estimation with complex data", IEEE Trans. Signal Process., Vol. 43, No. 8, pp. 2030–2033, Aug. 1995.

[26] P. J. Schreier, L. L. Scharf,"Second-order analysis of improper complex random vectors and process", IEEE Trans. Signal Process., Vol. 51, No. 3, pp. 714–725, Mar. 2003.

[27] S. Javidi, M. Pedzisz, S. L. Goh, D. Mandic,"The augmented complex least mean square algorithm with application to adaptive prediction problems", in Proc. 1st IARP Workshop Cognitive Inform. Process., 2008, pp. 54–57.

[28] S. L. Goh, D. P. Mandic,"An augmented extended Kalman filter algorithm for complex-valued recurrent neural networks", Neural Com-put., Vol. 19, No. 4, pp. 1039–1055, Apr. 2007.

[29] S. L. Goh, D. P. Mandic,"An augmented CRTRL for complex-valued recurrent neural networks", Neural Netw., Vol. 20, No. 10, pp. 1061–1066, 2007.

[30] B. Widrow, J. McCool, M. Ball,"The complex LMS algorithm", Proc. IEEE, Vol. 63, No. 4, pp. 719–720, Apr. 1975.

[31] P. Schreier, L. Scharf,"Statistical Signal Processing of Complex-Valued Data: The Theory of Improper and Noncircular Signals. Cam-bridge, U.K.: Cambridge Univ. Press, 2010.

ISSN : 0976-8491 (Online) | ISSN : 2229-4333 (Print)

Ms. V. V. Durga Bhavani received Masters in computer application from Acharya Nagarjuna University, Guntur in 2004. She worked as Lecturer, Asst. Professor and Associate professor & Head of MCA Department. At present, She is currently engaged in M.Tech, CSE from Jawharlal Nehru University. Her current research paper on Adaptive filtering of intricate Nonlinear Singnals.

Dr. Syed Sadat Ali Ph.D

Dr. Kishore Kumar, Ph.D