Chaotic Time Series Prediction Via Quaternionic Multilayer Perceptrons

P. Arena, S. Baglio, L. Fortuna, M. G. Xibilia
Dipartimento Elettrico Elettronico e Sistemistico
Universita' di Catania
Viale A. Doria 6, 95100 Catania, Italy
Tel: +39-95-339535 Fax: +39-95-330793
E-mail: xibi@dees.unict.it

ABSTRACT

In the paper a new type of Multi Layer Perceptron, developed in Quaternion algebra (QMLP), is adopted in order to predict chaotic time series. The use of QMLPs allows to perform accurate time series prediction with a decreased network complexity with respect to the classical real valued MLP, when the involved time series are multidimensional. The approach proposed has been adopted to estimate the chaotic behavior of Chua's circuit and of a circuit containing a piece-wise linear hysteresis element. A comparison between the performance of the QMLP and the real MLP is also reported in order to show the improvement introduced by the QMLP in terms of a sentivive decreasing of the network complexity.

1. INTRODUCTION

In the last few years artificial neural networks, and in particular Multi-Layer Perceptrons (MLP), have been introduced in order to predict non-linear time series as well as to perform continuous real valued function approximation and to carry on non linear system identification.

Dealing in particular with chaotic time series, the prediction accuracy decreases when the prediction time increases. In this field MLPs have shown better results with respect to conventional methods at large prediction time steps [1]. When the system taken into consideration involves signals defined in multidimensional domains, as often occurs in chaotic systems, a drawback can arise, caused from the growing of the network size. In particular, when the size of the input and output layer increases, the number of the weights connecting such layers to the hidden one heavily grows, thus slowing down the learning phase without contributing to the network generalization capabilities and enhancing the possibility of being trapped into local minima. In this perspective the proposed strategy is aimed to reduce the influence of such phenomenon, enhancing the generalization properties.

A new type of MLP developed in Quaternion Algebra (QMLP) is therefore introduced as a powerful strategy to decrease the number of free parameters , as shown in the application proposed. The introduced QMLP allows to perform complex non-linear system and time-series prediction employing a reduced complexity with respect to real MLP in order to reach the same or even better performance.

2. THE PROPOSED QMLP

One of the drawbacks of MLP structures is that the connections number heavily increases with the input/output dimension. In order to overcome this problem a new type of MLP developed in Quaternion Algebra [2] has been proposed (QMLP) in [3], [4], [5], where the suitable learning algorithm is derived. Quaternion Algebra has been introduced by Hamilton in 1843 to extend Complex Algebra in three dimensional spaces.

Quaternion Algebra

A quaternion $\bar{q}$ is defined as a complex number:

$$\bar{q} = q_o + q_1\bar{i} + q_2\bar{j} + q_3\bar{k}$$

formed from four different units $1, \bar{i}, \bar{j}, \bar{k}$ by means of the real parameters, q_i (i=0,..4), where $\bar{i}, \bar{j}, \bar{k}$ are the 3 orthogonal spatial vectors. It is convenient to represent $\bar{q}$ in the matrix form:

$$\bar{q} = [q_o, q_1, q_2, q_3]^T$$

The conjugate of a quaternion is denoted by $\bar{q}^*$ and is defined by:

$$\bar{q}^* = q_o - q_1\bar{i} - q_2\bar{j} - q_3\bar{k}$$

Addiction and subtraction of two quaternions $\bar{q}$ and $\bar{p}$ are defined, in the matrix form, as:

$$\bar{q} \pm \bar{p} = [q_0 \pm p_0, q_1 \pm p_1, q_2 \pm p_2, q_3 \pm p_3]^T$$

Quaternion multiplication is defined as:

$$\bar{q} \otimes \bar{p} = q_0 p_0 - \vec{p} \cdot \vec{q} + q_0 \vec{p} + p_0 \vec{q} + \vec{q} \times \vec{p}$$

where:

$$\vec{q} = [q_1, q_2, q_3]^T$$
$$\vec{p} = [p_1, p_2, p_3]^T$$

and • and × represent the usual scalar and vector product respectively.
It has to be remarked that quaternion multiplication is not commutative, so that the space of quaternions has the algebraic structure of a ring.
Unlike spatial vectors, the set of quaternions forms a division algebra since for each non zero quaternion there is an inverse such that:

$$\bar{q} \otimes \bar{q}^{-1} = \bar{q}^{-1} \otimes \bar{q} = 1.$$

The inverse is given by:

$$\bar{q}^{-1} = \frac{\bar{q}^*}{\bar{q}^* \otimes \bar{q}}$$

where:

$$\bar{q}^* \otimes \bar{q} = \|\bar{q}\|^2$$

is the norm of $\bar{q}$.

Quaternion MLP

Let us define a quaternion MLP (QMLP), a MLP in which both the weights of connections and the biases are quaternions, as well as input and output signals.
For this neural structure, the following notation will be adopted:
M :number of layers in the network;
l :layer index (l=0 and l=M denotes the input and output layer respectively);
N_l:number of neurons of the l-th layer;
n: neuron index;
$\bar{X}_n^l = X_{0n}^l + \bar{i}X_{1n}^l + \bar{j}X_{2n}^l + \bar{k}X_{3n}^l$:
quaternion output of the n-th neuron of the l-th layer;
(In particular $\bar{X}_0^l = 1$ represent the bias inputs, $\bar{X}_n^0$ n=1,.. N_0 is the input signal and $\bar{X}_n^M = \bar{Y}_n$ n=1,.., N_M is the output)
$\bar{w}_{nm}^l = w_{0nm}^l + \bar{i}w_{1nm}^l + \bar{j}w_{2nm}^l + \bar{k}w_{3nm}^l$:
quaternion sinaptic weigth of the n-th neuron of the l-th layer, relative to the m-th output of the (l-1)-th layer;
$\bar{S}_n^l = S_{0n}^l + \bar{i}S_{1n}^l + \bar{j}S_{2n}^l + \bar{k}S_{3n}^l$:
'net function' relative to the n-th neuron of the l-th layer;
$\bar{\theta}_n^l = \theta_{0n}^l + \bar{i}\theta_{1n}^l + \bar{j}\theta_{2n}^l + \bar{k}\theta_{3n}^l$:
bias of the n-th neuron of the l-th layer;
$\bar{t}_n = t_{0n} + \bar{i}t_{1n} + \bar{j}t_{2n} + \bar{k}t_{3n}$, n=1,..M: target of the n-th output of the network.
The forward phase of the QMPL is described by the following equations:

Forward phase:
for l=1,..,M and n=1,.. N_l

$$\bar{S}_n^l = \sum_{m=0}^{N_{l-1}} \bar{w}_{nm}^l \otimes \bar{X}_m^{l-1} + \bar{\theta}_n^l \qquad (1)$$

$$\bar{X}_n^l = \bar{\sigma}(\bar{S}_n^l) \qquad (2)$$

where the quaternion activation function $\bar{\sigma}(\bar{q})$ is defined by:

$$\bar{\sigma}(q_0 + \bar{i}q_1 + \bar{j}q_2 + \bar{k}q_3) = \sigma(q_0) + \bar{i}\,\sigma(q_1) + \bar{j}\sigma(q_2) + \bar{k}\,\sigma(q_3) \qquad (3)$$

and $\sigma(\cdot)$ is the classical real-valued sigmoidal function.
As in the real back-propagation algorithm [10], the learning procedure involves the presentation of a set of pair of samples, in the quaternion space, which represent the inputs and outputs of the network.
During the forward phase the network computes its output and compares it with the target.
An error function is then computed as:

$$E = \frac{1}{2}\sum_{n=1}^{N_M}(\bar{t}_{pn} - \bar{X}_{pn}^M)^2 \qquad (4)$$

where p indicates the p-th pattern.
The weights are updated at time $k+1$, back-propagating the error in such a way as to perform a steepest descent on a surface in the weight space whose heigth at any point is equal to the error measure;
In the following, the derived training algorithm is reported.

Learning phase:
for l=1,..,M and n=1,.. N_l

$$\bar{e}_n^l = \bar{t}_n - \bar{X}_n^M \quad \text{for } l = M$$

$$\bar{e}_n^l = \sum_{h=1}^{N_{l+1}} \bar{w}_{hn}^{l+1} \otimes \bar{\delta}_h^{l+1} \quad \text{for } l = M-1,..,1 \qquad (5)$$

where:

$$\bar{\delta}_n^l = e_{0n}^l\sigma(S_{0n}^l) + \bar{i}e_{1n}^l\sigma(S_{1n}^l) + \bar{j}e_{2n}^l\sigma(S_{2n}^l) + \bar{k}e_{3n}^l\sigma(S_{3n}^l)$$

$$\bar{w}_{nm}^l(k+1) = \bar{w}_{nm}^l(k) + \varepsilon\bar{\delta}_n^l \otimes \bar{X}_m^{*l-1}$$

$$m = 0,\ldots,N_{l-1} \qquad (6)$$

$$\bar{\theta}_n^l(k+1) = \bar{\theta}_n^l(k) + \varepsilon\bar{\delta}_n^l$$

and ε is the learning rate.
The approximation capabilities of QMLP have been investigated in [5] where a density theorem is proven. In particular it is stated that QMLP can approximate with the desired accuracy any quaternion-valued non-linear function defined in a compact subset of H.
This result is a generalization of the density theorem proved for MLP in the real space [6].

3. APPLICATIONS AND RESULTS

The approach proposed has been applied to perform time series prediction of two chaotic circuits. The performance of the neural network have been evaluated by computing the correlation function between the actual time series and the predicted one for each state-variable, over a set of testing samples never presented to the networks:

$$\rho_s(\tau)=\frac{\frac{1}{M}\sum_{t=1}^{M}[O_s(k)-O_{sm}][O_s^{'}(k)-O_{sm}^{'}]}{\sqrt{\frac{1}{M}\sum_{t=1}^{M}[O_s(k)-O_{sm}]^2}\sqrt{\frac{1}{M}\sum_{t=1}^{M}[O_s^{'}(k)-O_{sm}^{'}]^2}} \quad (7)$$

where O_{sm} is the mean value of the generic state variable O_s $s=t,x,y,z$ while $O_{sm}^{'}$ is the mean value of the τ step ahead estimated state varible $O_s^{'}$ and M is the number of testing patterns.

The correlation functions have been computed when the prediction step τ increases from 1 to 10. In order to select the optimal neural network topology the growing strategy has been adopted, it means that the number of hidden neurons has been increased until good performance have been reached, both for the real and the quaternionic MLPs.

Chaotic circuit with hysteresis element.

The time series considered describes the behavior of a chaotic circuit with four state-variable, including a dependent voltage source characterized by a piece-wise linear hysteresis function [7].

The circuit dynamics undergoes transition from torus doubling route to area and volume expanding chaos when the width of hysteresis threshold is modified.

The circuit dynamics is governed by following equation:

$$\dot{q}_0=-q_0-\eta p_1 h(z)+q_1-\eta(p_1/\beta_1)h(z)$$
$$\dot{q}_1=-\alpha_1(q_0-\eta p_1 h(z))+\alpha_1\beta_1(q_1-\eta(p_1/\beta_1)h(z))$$
$$\dot{q}_2=-q_2-\eta p_2 h(z)+q_3-\eta(p_2/\beta_2)h(z)$$
$$\dot{q}_3=-\alpha_2(q_2-\eta p_2 h(z))+\alpha_2\beta_2(q_3-\eta(p_2/\beta_2)h(z))$$

where:

$h(z)=1$ *for* $z\geq -1$ and $h(z)=1$ *for* $z\leq 1$

$$z=q_0+q_2$$
$$p_1=\beta_1/(1-\beta_1)$$
$$p_2=\beta_2/(1-\beta_2)$$

h is the normalized hysteresis and is switched from 1 to -1 if z hits the left threshold -1 and viceversa. This system has five parameters $(\alpha_1,\beta_1,\alpha_2,\beta_2,\eta)$. which control oscillation frequencies, dumping and the width of the hysteresis threshold respectively.

The corresponding time series has been derived by simulating the circuit equations with ΔT=0.002 and the following values of initial condition and parameters which lead to volume expanding chaos:

$$(q_0,q_1,q_2,q_3)=(1,0,1,0)$$

$$(a_1,b_1,a_2,b_2,\eta)=(7.5,0.16,15,0.097,1.3)$$

In order to train the neural networks a set of 500 pattern has been built, 250 have been used during the learning phase while the remaining have been used to test the network performance by computing the correlation functions for each state variable.

Several topologies have been trained both for the real and the quaternionic MLPs, in particular a number of hidden neurons growing from 1 to 4 for the QMLP and from 4 to 12 for the real MLP have been considered. The number of I/O neurons is 1 for QMLPs and 4 for the real MLPs. The best performance has been obtained with the 1-2-1 QMLP and the 4-8-4 real MLP. In Fig. 1a,b,c,d a comparison of the correlation trends obtained with these networks for each state-variable is reported.

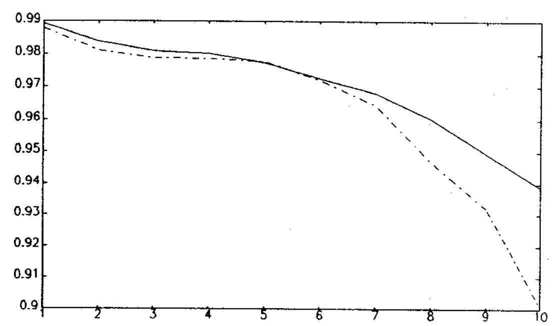

Fig 1

Comparison between the correlation trends of the 1-2-1 QMLP (-) and the 4-8-4 real MLP (-.) for the q_0 state variable

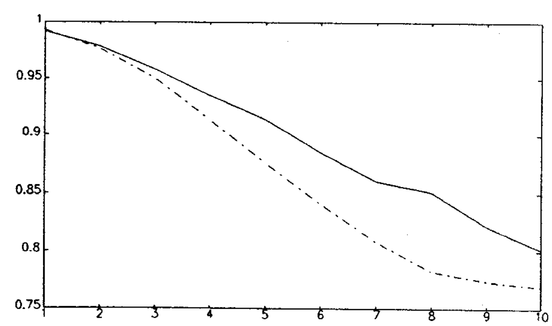

Fig 2

Comparison between the correlation trends of the 1-2-1 QMLP (-) and the 4-8-4 real MLP (-.) for the q_1 state variable.

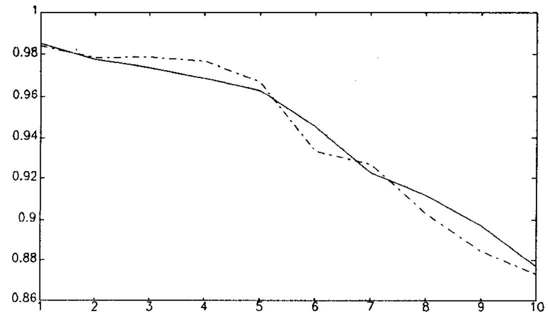

Fig 3

Comparison between the correlation trends of the 1-2-1 QMLP (-) and the 4-8-4 real MLP (-.) for the q_2 state variable.

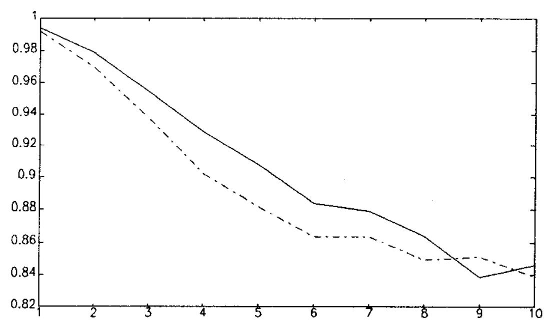

Fig 4

Comparison between the correlation trends of the 1-2-1 QMLP (-) and the 4-8-4 real MLP (-.) for the q_3 state variable.

As shown in the figures, the real and quaternionic MLPs give similar performance. However a strong difference can be observed when comparing the number of real parameters involved into the networks, which is 28 for the 1-2-1 QMLP and 76 for the 4-8-4 real MLP.

Chua's circuit

The well known Chua's circuit is the simplest circuit with the most complex behavior [8]. It is described by the following equations, here reported in the normalized form:

$$\dot{q}_1 = \alpha\left[y - h(q_1)\right]$$
$$\dot{q}_2 = q_1 - q_2 + q_3$$
$$\dot{q}_3 = -\beta\, q_2 - \gamma\, q_3$$
$$h(z) = m_1 q_1 + 0.5(m_0 - m_1)(|x+1| - |x-1|)$$

The 'double scroll' actractor is obtained for the following value of the parameters:

$$(\alpha,\beta,\gamma,m_0,m_1) = (9, 14.286, 0, -1/7, 2/7)$$

The initial conditions have been set to:

$$(x,y,z) = (0.1, 0.1, 0.1)$$

Starting from these conditions a set of 500 samples has been collected: among them 250 have been considered to perform the network learning and the remaining one for the testing step. In order to compare the real and the QMLP performance, the learning parameters as well as the maximum learning cycles equal to 10000 have been fixed to the same values for each topology.

In particular, as regards the QMLP, topologies employing from 1 to 5 hidden units have been taken into considerations. For such networks the real part of the input and output neurons has been set to zero. The best results have been reached with the QMLP with 3 hidden units, which employes 40 real parameters. As regards the real MLP, topologies starting from 6 to 12 hidden units have been trained. After performing such step, it has been noted that none of the topologies reaches a value of the testing error comparable or lower than the 1-3-1 QMLP testing error within the fixed cycle number. The comparison between the performance of the different topologies has been performed using the correlation function (7) evalued with a prediction step increased from 1 to 10. Fig. 2a,b,c clearly show the efficiency of the QMLP structures with respect to the real ones.

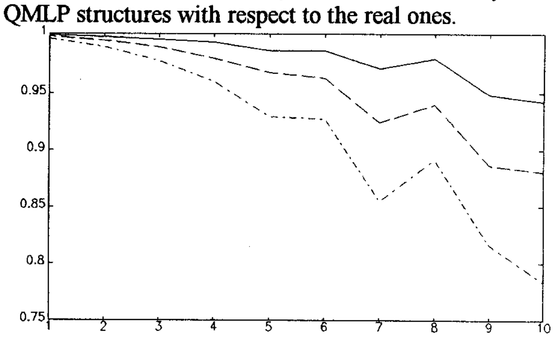

Fig. 5

Comparison between the correlation trends of the 1-3-1 QMLP after 10000 lerning cycles (-), the 3-12-3 real MLP after 10000 cycles (-.) and the 3-12-3 real MLP after 20000 cycles (--) for the q_1 state variable.

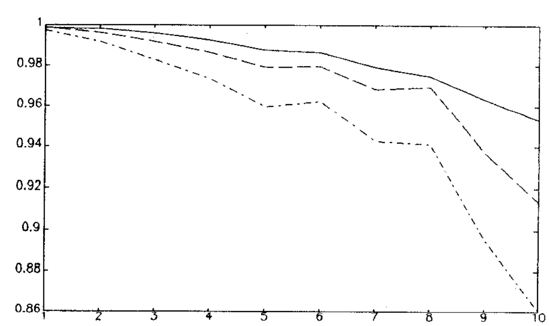

Fig. 6

Comparison between the correlation trends of the 1-3-1 QMLP after 10000 lerning cycles (-), the 3-12-3 real MLP after 10000 cycles (-.) and the 3-12-3 real MLP after 20000 cycles (--) for the q_2 state variable.

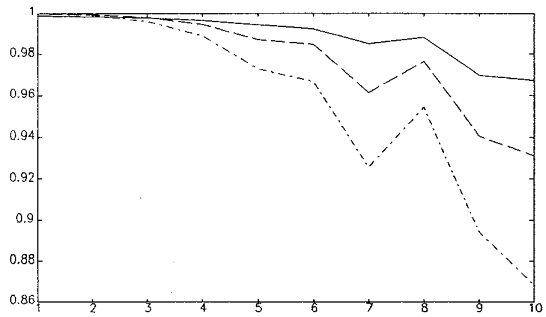

Fig. 7

Comparison between the correlation trends of the 1-3-1 QMLP after 10000 lerning cycles (-), the 3-12-3 real MLP after 10000 cycles (-.) and the 3-12-3 real MLP after 20000 cycles (--) for the q_3 state variable.

It has to be observed that the real MLP with 12 hidden units improves its performance when the number of learning cycles is increased but in any case it employs a number of real parameters equal to 87, more than twice with respect to the QMLP. Under these consideration the suitability of the proposed strategy acquires more interest.

4. CONCLUSIONS

In the paper a strategy in order to decrease the number of parameters needed to perform a multidimensional time series prediction has been introduced. Such a methodology makes use of a Multi Layer Perceptron defined in Quaternion Algebra, for which a suitable density theorem has been proven. The efficiency of the proposed approach has been validated through a series of examples regarding the estimation of chaotic time series derived from some chaotic circuits well known in literature.

5. REFERENCES

[1] A. Lapedes, R. Farber, "How neural nets work", Proc. of IEE Denver Conf. on Neural Nets, 1987.

[2] W. R. Hamilton, "Elements of Quaternions", 3rd edition, Chelsea, New York, 1969.

[3] P. Arena, L. Fortuna, L. Occhipinti, M.G. Xibilia, "Neural networks for quaternion-valued function approximation" IEEE Proc. of ISCAS '94, London, vol. 6.

[4] P. Arena, L. Fortuna, M.G. Xibilia, "Neural Networks in Multidimensional Domains" , accepted for publication in Journal of System Eng., U.K.

[5] M. G. Xibilia, "Algoritmi di apprendimento in algebra ipercomplessa e predizione di comportamenti caotici in circuiti e sistemi dinamici", Ph. D. Thesis, University of Catania, Italy, 1995.

[6] G. Cybenko, "Approximation by superposition of sigmoidal functions", Mathematics of control signal and systems, vol 2, 1989.

[7] K.Mitsubori, T. Saito, "Torus Doubling and Hyperchaos in a Five Dimensional Hysteresis Circuit", IEEE Proc. of ISCAS '94, London, vol. 6.

[8] R.N. Madan, "Chua's circuit. A paradigm for chaos", World Scientific Publ. Co., Singapore '93.