Proceedings of the 17th World Congress
The International Federation of Automatic Control
Seoul, Korea, July 6-11, 2008

Feedback Analysis of Radial Basis Functions Neural Networks via Small Gain Theorem ⋆

Ali, S. Saad Azhar * **Muhammad Shafiq** ** **Jamil M. Bakhashwain** *** **Fouad M. AL-Sunni** ****

* *Electrical Engineering Department, Air University, E-9 Islamabad, Pakistan, (email: saadali@mail.au.edu.pk).*
** *Department of Electronics Engineering, Ghulam Ishaq Khan Institute, Topi, Pakistan, (email: mshafiq@giki.edu.pk).*
*** *Electrical Engineering Department, King Fahd University of Petroleum & Minerals, Dhahran 31261, Saudi Arabia, (email: jamilb@kfupm.edu.sa).*
**** *Systems Engineering Department, King Fahd University of Petroleum & Minerals, Dhahran 31261, Saudi Arabia, (email: alsunni@kfupm.edu.sa).*

Abstract: Radial basis function neural networks are used in a variety of applications such as pattern recognition, nonlinear identification, control, time series prediction, etc. In this paper, feedback analysis of the learning algorithm of radial basis function neural networks is presented. It studies the robustness of the learning algorithm in the presence of uncertainties that might be due to noisy perturbations at the input or to modeling mismatch. The learning scheme is first associated with a feedback structure and then the stability of that feedback structure is analyzed via small gain theorem. The analysis suggests bounds on the learning rate in order to guarantee that the learning algorithm will behave as robust nonlinear filters and optimal choices for faster convergence speeds.

1. INTRODUCTION

Neural networks have been recently used widely in a variety of areas such as pattern recognition, system identification, filtering, control, time series prediction, etc. Radial basis function neural networks (RBFNN) are single-layered feedforward networks with universal approximation capabilities, in addition to more efficient learning than the famous multi-layered feedforward neural networks (MFNN) Haykin [1999], Jun-Dong et al. [1998], Finan et al. [1996], Fortuna et al. [2001].

RBFNN are generally trained using supervised learning. During training, a recursive update procedure is used to estimate the weights of the RBFNN that best fits the given data Haykin [1999]. The recursive procedure often requires to select a suitable adaptation gain called learning rate. The learning rate should be within an optimum range. It should neither be too large which would drive the algorithm unstable, nor too small, that it slows down the training. In general practice, trial-and-error experiences are used to select a suitable learning rate for training phase.

The general and simpler practice has been to choose a small learning rate that obviously result in slower convergence speeds. Especially, with multivariable systems with many weights and a large data, a small learning rate may require substantial amount of time and machine power.

Therefore, it should be analyzed to find an optimal learning rate to speed up the convergence and yet keeping the algorithm stable. In the robustness analysis of adaptive schemes Sayed et al. [1996] and Rupp et al. [1995], the authors have addressed the methods of selecting the learning rate 1) in order to guarantee a robust behavior in the presence of noise and modeling uncertainties and 2) in order to guarantee a faster convergence speeds.

The formulation in Sayed et al. [1996] and Rupp et al. [1995] emphasizes an intrinsic feedback structure for most adaptive algorithms and it relies on tools from system theory, control and signal processing such as state-space description, feedback analysis, small gain theorem, H^{∞} design and lossless systems. The feedback configuration is provoked via energy arguments and is shown to consist of two major blocks: a time-variant *lossless* (*i.e.*, energy preserving) feedforward path and a time-variant feedback path.

We make use of the feedback structure to analyze robustness of RBFNN and find optimal choices for learning rate. In this paper, we present the learning algorithm for the RBFNN, that involves a nonlinear functional in the update equation due to the presence of the basis function (usually a gaussian function) and associate with the feedback structure of Sayed et al. [1996] and Rupp et al. [1995] in order to handle the presence of the nonlinearity. As an

⋆ This work is sponsored by King Fahd University of Petroleum & Minerals and SABIC under project SABIC 2006-11

argument, we suggest the choice for the learning rate in order to guarantee stability, faster convergence and robust performance.

This paper is organized in five sections. In section 2 robustness issues are discussed using a contractive mapping and bounds for the learning rate are suggested. Section 3 associates the learning algorithm with the feedback structure, optimal choice for learning rate via small gain theorem is presented in section 4, and the paper is concluded in section 5.

1.1 Radial Basis Functions Neural Networks

RBFNN is a type of feedforward neural network. They are used in a wide variety of contexts such as function approximation, pattern recognition and time series prediction. Networks of this type have the universal approximation property Haykin [1999]. In these networks the learning involves only one layer with lesser computations. A single input single output RBFNN is shown in Fig. 1. The RBFNN consists of an input node $u(t)$, a hidden layer with n_o neurons and an output node $y(t)$. Each of the input node is connected to all the nodes or neurons in the hidden layer through unity weights (direct connections). While each of the hidden layer nodes is connected to the output node through some weights, *e.g.* the i^{th} output node is connected with all the hidden layer nodes by $W(t) = [w_1(t), \ldots, w_{n_o}(t)]$. Each neuron finds the distance, normally applying Euclidean norm, between the input and its center and passes the resulting scalar through a non-linearity. So the output of the i^{th} hidden neuron is given by $\phi_i(\|u(t) - c_i\|)$, where c_i is the center of the i^{th} hidden layer node, $i = 1, 2, \ldots, n_o$, and $\phi_i(\cdot)$ is the nonlinear basis function. Normally this function is taken as a Gaussian function of width β, that dictates the effective range of input passing through the basis function. Normally, β should at least be equal to the spacing between the neurons. The output $y_m(t)$ is a weighted sum of the outputs of the hidden layer, given by

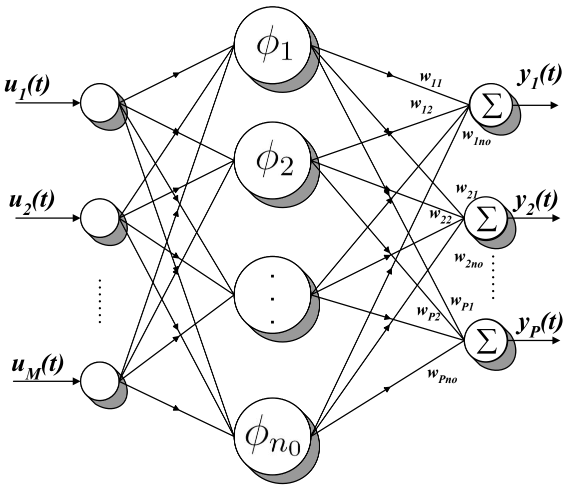

Fig. 1. A MIMO RBF neural network.

$$y_m(t) = \Phi(t)W(t), \tag{1}$$

$$y_m(t) = \sum_{i=1}^{n_o} \phi_i(\|u(t) - c_i\|)w_i(t),$$

where the basis functions and weight vector are defined as,

$$\Phi(t) = [\phi_1(u(t))\ \phi_2(u(t))\ \ldots\ \phi_{n_o}(u(t))], \tag{2}$$

$$W(t) = [w_1(t)\ w_2(t)\ \ldots\ w_{n_o}(t)]^T. \tag{3}$$

and the gaussian basis function is,

$$\phi_i(u(t)) = exp\big(-\frac{\|u(t) - C_i)\|^2}{\beta^2}\big) \tag{4}$$

Consider a collection of input vectors $\{u(t)\}$ with the corresponding desired output vectors $\{y(t)\}$. We also take into account noisy perturbations $v(t)$ in the desired signal. These perturbations can be due to model mismatch or to measurement noise. Assuming there exists an optimal weight vector W^* such that

$$y(t) = \Phi(t)W^* + v(t). \tag{5}$$

The RBFNN is presented with the given input-output data $\{u(t), y(t)\}$. The objective is to estimate the unknown optimal weight W^*. Now, starting with an initial guess W_0, the weights are updated recursively based on the LMS principle as,

$$W(t+1) = W(t) + \alpha(t)e(t)\Phi^T(t) \tag{6}$$

where $\alpha(t)$ is the learning and the error $e(t)$ is defined as,

$$e(t) = y(t) - y_m(t) + v(t)$$
$$e(t) = \Phi(t)W^* - \Phi(t)W(t) + v(t) \tag{7}$$

Defining *a priori* and *a posteriori* error quantities as

$$e_a(t) = \Phi(t)\tilde{W}(t) \tag{8}$$

$$e_p(t) = \Phi(t)\tilde{W}(t+1) \tag{9}$$

where $\tilde{W}(t)$ is the weight error vector symbolizing the difference between the optimal weight and its estimate as $\tilde{W}(t) = W^* - W(t)$. Therefore,

$$\begin{aligned} e_a(t) &= \Phi(t)(W^* - W(t)), \\ &= \Phi(t)W^* - \Phi(t), W(t) \\ &= \Phi(t)W^* - y_m(t). \end{aligned}$$

and the weight error update equation satisfies the following recursion,

$$\tilde{W}(t+1) = \tilde{W}(t) + \alpha(t)e(t)\Phi^T(t) \tag{10}$$

2. ROBUSTNESS

It is imperative to bring up that a robust algorithm has consistent estimation errors with the disturbances in the sense that "small" disturbances would lead to "small" estimation errors, no matter what the disturbances are. Generally, this is not the case for any adaptive algorithm. The estimation errors can still be large even in the presence of small disturbances Hasibi et al. [1996].

The robustness issue is dealt here in a purely deterministic framework and without assuming prior knowledge of signal or noise statistics. This is especially useful in situations where prior statistical information is missing. The robust design would guarantee a desired level of robustness independent of the noise statistics. In a broad sense, robustness would imply that the ratio of an estimation error energy to the noise or disturbance energy will be guaranteed to be upper bounded by a positive constant

$$\frac{estimation\ error\ energy}{disturbance\ energy} \leq 1 \tag{11}$$

As a matter-of-fact the approach given by the ratio in Eq. 11 is quiet desirable since it assures that the resulting estimation error energy will be upper bounded by the disturbance energy, regardless of the nature and statistics of noise. In the following section, the robustness methodology will be adopted to select the learning rate in order to guarantee robust behavior.

2.1 Optimal Learning Rate for Robustness

In this continuation, we will develop a contractive mapping from the t^{th} instant to $t+1^{th}$ instant of the recursion. A linear map that transforms x to y, as $y = T[x]$, is said to be contractive mapping, if for all x we have $\|T[x]\|^2 \leq \|x\|^2$. This depicts that the output energy does not exceed the input energy. The contractive mapping will relate the energies in such a way that the ratio in Eq. 11 is satisfied. More specifically, the Euclidean norm of the weight error vector and the *a priori* estimation errors at the $t+1^{th}$ instant is compared with and the Euclidean norms of the weight error vectors and disturbance error.

The disturbance error can be defined as,

$$\tilde{v}(t) = e(t) - e_a(t) \tag{12}$$

Now consider the weight error recursion given by Eq. 10

$$\tilde{W}(t+1) = \tilde{W}(t) - \alpha(t)e(t)\Phi^T(t).$$

The squared norm in effect the energies, of the weight error recursion equation can be computed as follows,

$$\begin{aligned}
\|\tilde{W}(t+1)\|^2 &= \|\tilde{W}(t)\|^2 - 2\alpha(t)e(t)\Phi(t)\tilde{W}(t) \\
&\quad + \alpha(t)^2 e^2(t)\|\Phi(t)\|^2, \\
&= \|\tilde{W}(t)\|^2 - 2\alpha(t)\Phi(t)\tilde{W}(t)(e_a(t) + \tilde{v}(t)) \\
&\quad + \alpha(t)^2\|\Phi(t)\|^2(e_a(t) + \tilde{v}(t))^2, \\
&= \|\tilde{W}(t)\|^2 - 2\alpha(t)(e_a^2(t) + e_a(t)\tilde{v}(t)) + \\
&\quad \alpha(t)^2\|\Phi(t)\|^2(e_a^2(t) + 2e_a(t)\tilde{v}(t) + \tilde{v}^2(t)), \\
&= \|\tilde{W}(t)\|^2 - 2\alpha(t)e_a^2(t) - 2\alpha(t)e_a(t)\tilde{v}(t) + \\
&\quad \alpha(t)^2\|\Phi(t)\|^2 e_a^2(t) + 2\alpha(t)^2\|\Phi(t)\|^2 e_a(t)\tilde{v}(t) \\
&\quad + \alpha(t)^2\|\Phi(t)\|^2\tilde{v}^2(t).
\end{aligned}$$

Rearranging terms we get,

$$\begin{aligned}
&\|\tilde{W}(t+1)\|^2 + 2\alpha(t)e_a^2(t) - \alpha(t)^2\|\Phi(t)\|^2 e_a^2(t) = \\
&\quad \|\tilde{W}(t)\|^2 - 2\alpha(t)e_a(t)\tilde{v}(t) + 2\alpha(t)^2\|\Phi(t)\|^2 e_a(t)\tilde{v}(t) \\
&\quad + \alpha(t)^2\|\Phi(t)\|^2\tilde{v}^2(t).
\end{aligned} \tag{13}$$

Introducing a parameter $\mu(t)$ as

$$\mu(t) = \frac{1}{\|\Phi(t)\|^2} \tag{14}$$

Using $\mu(t)$ in Eq. 13, we get

$$\|\tilde{W}(t+1)\|^2 + 2\alpha(t)e_a^2(t) - \frac{\alpha(t)^2}{\mu(t)}e_a^2(t) =$$

$$\|\tilde{W}(t)\|^2 - 2\alpha(t)e_a(t)\tilde{v}(t) + 2\frac{\alpha(t)^2}{\mu(t)}e_a(t)\tilde{v}(t) + \frac{\alpha(t)^2}{\mu(t)}\tilde{v}^2(t).$$

If we set $\alpha(t) = \mu(t)$, we come up to the following equality, where the energy bounds are always satisfied as estimation energy = disturbance energy.

$$\begin{aligned}
&\|\tilde{W}(t+1)\|^2 + 2\alpha(t)e_a^2(t) - \alpha(t)e_a^2(t) = \\
&\|\tilde{W}(t)\|^2 - 2\alpha(t)e_a(t)\tilde{v}(t) + 2\alpha(t)e_a(t)\tilde{v}(t) + \alpha(t)\tilde{v}^2(t) \\
&\quad \|\tilde{W}(t)\|^2 + \alpha(t)e_a^2(t) = \|\tilde{W}(t)\|^2 + \alpha(t)\tilde{v}^2(t).
\end{aligned}$$

Therefore, we can conclude to the results for the energy bounds depending upon the learning rate.

$$\frac{\|\tilde{W}(t+1)\|^2 + \alpha(t)e_a^2(t)}{\|\tilde{W}(t)\|^2 + \alpha(t)\tilde{v}^2(t)} \begin{cases} \leq 1 \ for\ 0 < \alpha(t) < \mu(t) \\ = 1 \ for\ \alpha(t) = \mu(t) \\ \geq 1 \ for\ \alpha(t) > \mu(t) \end{cases} \tag{15}$$

The first two inequalities in the statement of Eq. 15 ascertain that if the learning rate is chosen such that $\alpha(t) \leq \mu(t)$, then the mapping from from the signals $\{\tilde{W}(t), \sqrt{\mu(t)}e_p(t)\}$ to the signals $\{\tilde{W}(t+1), \sqrt{\mu(t)}e_a(t)\}$ is a contractive mapping. Therefore, a local energy bound is deduced that highlights a robustness property of the update recursion. The energy bound depict that no matter what the value of the noise component $\tilde{v}(t)$ is, and no matter how far the estimate $W(t)$ is from the optimal W^*, the sum of energies $\|\tilde{W}(t+1)\|^2 + \alpha(t)e_a^2(t)$ will always be smaller than or equal to the sum of energies $\|\tilde{W}(t)\|^2 + \alpha(t)\tilde{v}^2(t)$.

Remarks: Since this contractivity property holds for each t^{th} instant, it should also hold globally over any interval. In fact, selecting $\mu(t) < \gamma(t)$ over the interval $0 \leq t \leq N$, it follows that,

$$\|\tilde{W}_N\|^2 + \sum_{t=0}^{N}\alpha(t)e_a^2(t) = \|\tilde{W}_0\|^2 + \sum_{t=0}^{N}\alpha(t)\tilde{v}^2(t).$$

3. FEEDBACK STRUCTURE

The bounds of the statement given by 15 can be illustrated in an alternative structure that establish the feedback structure. Initially, the recursive weight update equation has to be written as a function of *a priori error* and *a posteriori error*.

The *a posteriori error* s defined in 9 as,

$$\begin{aligned}
e_p(t) &= \Phi(t)\tilde{W}(t+1) \\
&= \Phi(t)[\tilde{W}(t) - \alpha(t)\Phi(t)^T e(t)] \\
&= e_a(t) - \alpha(t)\|\Phi(t)\|^2 e(t) \\
&= e_a(t) - \frac{\alpha(t)}{\mu(t)}e(t)
\end{aligned} \tag{16}$$

$$\mu(t)e_p(t) = \mu(t)e_a(t) - \alpha(t)e(t)$$

$$\mu(t)\big(e_a(t) - e_p(t)\big) = \alpha(t)e(t).$$

Hence, the recursive weight update Eq. 6 can be written as

$$W(t+1) = W(t) + \mu(t)\big(e_a(t) - e_p(t)\big)\Phi^T(t).$$

Similarly, the weight error recursive Eq. 10 can be reformulated as,

$$\tilde{W}(t) = \tilde{W}(t) - \mu(t)\Phi^T(t)\big(e_a(t) - e_p(t)\big) \qquad (17)$$

The squared norm of Eq. 17 leads to the same statement as 15, except that the disturbance error $\tilde{v}(t)$ is replaced by the negative of *a posteriori error* $-e_p(t)$ and the learning rate is set to $\mu(t)$.

$$\begin{aligned}
\|\tilde{W}(t+1)\|^2 &= \|\tilde{W}(t)\|^2 - 2\mu(t)\Phi(t)\tilde{W}(t)\big(e_a(t) - e_p(t)\big) \\
&\quad + \mu(t)^2\|\Phi(t)\|^2\big(e_a(t) - e_p(t)\big)^2 \\
&= \|\tilde{W}(t)\|^2 - 2\mu(t)e_a(t)\big(e_a(t) - e_p(t)\big) \\
&\quad + \mu(t)^2\frac{1}{\mu(t)}e_a^2(t) - 2\mu(t)^2\frac{1}{\mu(t)}e_a(t)e_p(t) \\
&\quad + \mu(t)^2\frac{1}{\mu(t)}e_p^2(t) \\
&= \|\tilde{W}(t)\|^2 - 2\mu(t)e_a^2(t) + 2\mu(t)e_a(t)e_p(t) \\
&\quad + \mu(t)e_a^2(t) - 2\mu(t)^2e_a(t)e_p(t) + \mu(t)^2e_p^2(t)
\end{aligned}$$

$$\|\tilde{W}(t+1)\|^2 + \mu(t)e_a^2(t) = \|\tilde{W}(t)\|^2 + \mu(t)e_p^2(t) \qquad (18)$$

$$\frac{\|\tilde{W}(t+1)\|^2 + \mu(t)e_a^2(t)}{\|\tilde{W}(t)\|^2 + \mu(t)e_p^2(t)} = 1. \qquad (19)$$

Hence, the energy ratio in Eq. 19 holds for all possible choices of the learning rate. This implies that the mapping $\overline{T}_i$ from the signals $\{\tilde{W}(t), \sqrt{\mu(t)}e_p(t)\}$ to the signals $\{\tilde{W}(t+1), \sqrt{\mu(t)}e_a(t)\}$ is lossless.

Now if we further apply the mean-value theorem to the output of the RBFNN $\Phi(t)W(t)$, we can write

$$\Phi(t)W^* - \Phi(t)W(t) = \Phi'(\tau)W(t)e_a(t) \qquad (20)$$

for some point τ along the segment connecting $\Phi(t)W^*$ and $\Phi(t)W(t)$. Therefore, combining Eq. 7 and Eq. 16,

$$\begin{aligned}
e_p(t) &= e_a(t) - \frac{\alpha(t)}{\mu(t)}e(t) \\
e_p(t) &= e_a(t) - \frac{\alpha(t)}{\mu(t)}\big(\Phi(t)W^* - \Phi(t)W(t) + v(t)\big) \\
e_p(t) &= e_a(t) - \frac{\alpha(t)}{\mu(t)}\big(\Phi'(\tau)W^*e_a(t) + v(t)\big) \\
e_p(t) &= [1 - \frac{\alpha(t)}{\mu(t)}\Phi'(\tau)W^*]e_a(t) - \frac{\alpha(t)}{\mu(t)}v(t) \\
-\sqrt{\mu(t)}e_p(t) &= \frac{\alpha(t)}{\sqrt{\mu(t)}}v(t) - [1 - \frac{\alpha(t)}{\mu(t)}\Phi'(\tau)W^*]\sqrt{\mu(t)}e_a(t)
\end{aligned} \qquad (21)$$

This relation shows that the overall mapping from the *original* (weighted) disturbances $\sqrt{\mu(t)}v(t)$ to the resulting *a priori* (weighted) estimation errors $\sqrt{\mu(t)}e_a(t)$ can be expressed in terms of a feedback structure, as shown in Fig. 2.

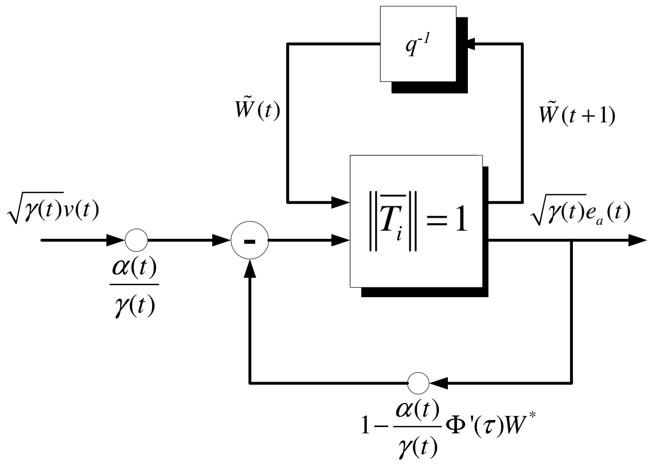

Fig. 2. A lossless mapping in a feedback structure for RBFNN learning algorithm

The stability of such structures can be studied via tools that are by now standard in system theory (*e.g.* the small gain theorem). Conditions on the learning rate $\alpha(t)$ will be derived in order to guarantee a robust training algorithm, as well as faster convergence speeds.

This will be achieved by establishing conditions under which the feedback configuration is l_2 stable in the sense that it should map a finite-energy input noise sequence (which include the noiseless case a special case) $\{\sqrt{\mu(t)}v(t)\}$ to a finite-energy *a priori* error sequence $\{\sqrt{\mu(t)}e_a(t)\}$

4. OPTIMAL LEARNING RATE VIA SMALL GAIN THEOREM

In order to make use of the tools from the system theory, such as the l_2 stability and small gain theorem, define

$$\gamma(N) = \max_{0 \le t \le N} \quad \frac{\alpha(t)}{\mu(t)} \quad , \qquad (22)$$

$$\Delta(N) = \max_{0 \le t \le N} \quad |1 - \frac{\alpha(t)}{\mu(t)}\Phi'(t)W^*| \quad . \qquad (23)$$

According to the definition in Eq. 23, $\Delta(N)$ is the maximum absolute gain of the feedback loop over the interval $0 \le t \le N$.

The small gain theorem states that the l_2 stability of a feedback configuration such as the configuration in Fig. 2 as special case requires that the product of norms of the feedforward and feedback maps be strictly strictly bounded by one.

In our case, the norm of the feedforward map is equal to one (since it is lossless) while the norm of the feedback map is defined in Eq. 23 as $\Delta(N)$. Hence, the condition $\Delta(N) < 1$ guarantees an overall contractive map.

Therefore, for $\Delta(N) < 1$ to hold, we need to choose the learning rate such that, for all t

$$0 < \alpha(t)\Phi'(\tau)W^* < 2\mu(t) = \frac{2}{\|u(t)\|^2} \qquad (24)$$

In Rupp et al. [1997], the authors have presented a number of choices for learning rate. They based the selection of learning rate on the availability of the derivative function $\Phi'(\tau)W^*$. For the case of RBFNN it is straight forward to obtain the estimate of the derivative function using the current basis function as,

$$\begin{aligned}
\Phi(\tau)W(\tau) &= [\phi_1(\tau) \ \dots \ \phi_{n_o}(\tau)]W(\tau),\\
&= [\exp(-\frac{\|u(\tau)-c_1\|^2}{\beta^2}) \ \dots \\
&\qquad \exp(-\frac{\|u(\tau)-c_{n_o}\|^2}{\beta^2})]W(\tau),\\
&= [\exp(-\frac{\|c_1-u(\tau)\|^2}{\beta^2}) \ \dots \\
&\qquad \exp(-\frac{\|c_{n_o}-u(\tau)\|^2}{\beta^2})]W(\tau),\\
\Phi'(\tau)W(\tau) &= -[\phi_1(\tau)\frac{\partial}{\partial u(\tau)}\frac{\|c_1-u(\tau)\|^2}{\beta^2} \ \dots \\
&\qquad \phi_{n_o}(\tau)\frac{\partial}{\partial u(\tau)}\frac{\|c_{n_o}-u(\tau)\|^2}{\beta^2}]W(\tau),\\
\Phi'(\tau)W(\tau) &= \frac{2}{\beta^2}[(c_1-u(\tau))\phi_1(\tau) \ \dots \\
&\qquad (c_{n_o}-u(\tau))\phi_{n_o}(\tau)]W(\tau). \qquad (25)
\end{aligned}$$

Defining $\overline{\Phi}(\tau)$ as $\overline{\Phi} = [(c_1 - u(\tau))\phi_1(\tau) \ \dots \ (c_{n_o} - u(\tau))\phi_{n_o}(\tau)]$, we get the derivative as,

$$\Phi'(\tau)W(\tau) = \frac{2}{\beta^2}\overline{\Phi}(\tau)W(\tau). \qquad (26)$$

Therefore, the derivative in Eq. 26 can be used to find the optimal learning rate to speed up the convergence as,

$$\alpha(t) < 2\mu(t)\Phi'(\tau)W(\tau), \qquad (27)$$

$$\alpha(t) < 2\mu(t)\frac{\beta^2}{2\overline{\Phi}(\tau)W(\tau)}, \qquad (28)$$

$$\alpha(t) < \mu(t)\frac{\beta^2}{\overline{\Phi}(\tau)W(\tau)}. \qquad (29)$$

Remarks: This optimal learning rate not only guarantees the stability of the feedback structure, *i.e.* the stability of the learning of RBFNN, but also ensures faster convergence speeds.

5. CONCLUSIONS

In this paper a feedback analysis of the learning algorithm of RBFNN is presented. The stability of the learning algorithm is analyzed using the small gain theorem by associating the algorithm with a feedback structure. Choices for suitable learning rates are suggested that guarantee a robust behavior in the presence of noise. In order to speed up the convergence, bounds for the optimal learning rate are presented.

ACKNOWLEDGEMENTS

The authors acknowledge the support of King Fahd University of Petroleum & Minerals and SABIC for funding this work under project SABIC 2006-11.

REFERENCES

Haykin, S. Neural Networks:A Comprehensive Foundation II. *Macmillan/IEEE Press*,1999.

Jun-Dong, P. and Jones, W. L. Comparison of neural network classifiers for NSCAT sea ice flag. *IEEE International Geoscience and remote sensing symposium proceedings,* ,pages 2237-2239, 1998.

Fortuna, L., Muscato, G. and Xibilia, M. G. A comparison between HMLP and HRBF for attitude control. *IEEE transactions on neural networks,* Volume 122, pages 318-328, March 2001.

Finan, R. A., Sapeluk, A. T. and Damper, R. I. Comparison of multilayer and radial basis functions neural networks for the text dependent speaker recognition. *IEEE International conference on Neural Networks,* Volume 4, pages 1992-1997, 1996.

Ali H. Sayed and M. Rupp. A time-domain feedback analysis of filtered-error adaptive algorithms via the samll gain theorem. *Proceedings of SPIE,*, Volume 2563, pages .458-469,June 1996.

M Rupp and Ali H. Sayed. A time-domain feedback analysis of filtered-error adaptive gradient algorithms. *IEEE Transactions on Signal Processing*,Volume 44, pages .1428-1439, July 1995.

Hasibi, B., Ali H. Sayed and Kailath, T. H^{∞} optimality of the LMS algorithm. *IEEE Transactions on Signal Processing*, Volume 44, pages -267-280, Feb. 1996.

Rupp, M. and Ali H. Sayed. Supervised learning of perceptron and output feedback dynamic networks: A feedback analysis via the small gain theorem. *IEEE Transactions on Neural Networks*, Volume 8, pages 612-622, May 1997.

Ali, S. Saad Azhar, Fouad M. AL-Sunni and Shafiq, M. U-model based adaptive tracking of unknown MIMO Bilinear Systems. *1st IEEE conference on industrial electronics and applications*, Singapore, May 2006.

Ali, S. Saad Azhar, Fouad M. AL-Sunni, Shafiq, M. and Jamil M. Bakhashwain. Learning Feedforward control of MIMO nonlinear systems using U-model. *9th IASTED conference on Control and Applications,* Montreal, Quebec, May 2007.

Shafiq, M. and Haseebuddin, M. Internal model control for nonlinear dynamic plants using U-model. *12th Mediterranean Conference on Control and Automation*, Turkey, June 2004.

Nabil R. Yousef and Ali H. Sayed. A unified approach to the steady-state and tracking analyses of Adaptive filters. *IEEE Transactions on Signal Processing*,Volume 49, no.2, pages .314-324,February 2001 .

Ali H. Sayed and M. Rupp. DSP Handbook. *Boca Raton*,FL:CRC, 1998, Ch.20.

E. Ewada. Comparison of RLS, LMS and sign algorithm for tracking of randomly time-varying channels. *IEEE Transactions on Signal Processing*,Volume 42, pages .2937-2944,Nov, 1994.