IASTED International Conference on
Intelligent Systems and Control (*ISC 2001*), *Nov. 19-22, Tampa, USA, 2001.*

State-Space Neural Networks and the Unscented Kalman Filter in On-line Nonlinear System Identification

P. Gil[†‡] pgil@dei.uc.pt
J.Henriques[†] jh@dei.uc.pt
H. Duarte-Ramos[‡] hdr@mail.fct.unl.pt
A. Dourado[†] dourado@dei.uc.pt

[†]CISUC - Informatics Engineering Department, UC
Pólo II, Pinhal de Marrocos 3030 Coimbra – Portugal
Phone: +351 239790000, Fax: +351 239701266

[‡]Electrical Engineering Department, UNL
2825 Monte de Caparica – Portugal
Phone: +351 212948545, Fax: +351 212948532

ABSTRACT

This paper focuses on online non-linear system identification via state-space neural networks. The training algorithm is based on a generalisation of the Kalman filter to non-linear systems by means of the unscented transformation. Experimental results from a laboratory heating system confirm the feasibility and effectiveness of the proposed methodology.

KEY WORDS

On-line identification, state-space neural networks, non-linear systems, the unscented Kalman filter.

1. Introduction

The problem of finding a suitable model structure and a particular parameterisation in order to describe the input-output behaviour of a given unknown system is the realm of system identification.

System's modelling and identification dates back to Karl F. Gauss (1777-1855), who first formulated the principle of least squares and applied it to determine the orbits of planets. Nevertheless, it was only around 1960 with the advent of modern control theory and the need for parametric models that extensive research in the field began [1]. In the last decades, prompted by the fact that most dynamical systems can be better characterised by non-linear models, which are theoretically able to describe the system's behaviour over the whole operating range, non-linear black-box modelling has attracted a great deal of interest, particularly among the control community. One class of these models that has been the subject of extensive research activity is that of neural networks [2], [3], [4], [5].

Despite purely feedforward neural networks under the form of tapped-delay-line representations have long been used to process temporal information, it is recognized that dynamic neural structures containing a state feedback may provide computational advantages: the corresponding non-linear state-space models are likely to possess a smaller number of parameters and, in addition, they make possible to describe a larger class of dynamic systems than tap-delay-line temporal representations [6]. Other factors, such as the quality of the training data set and the training methodology itself, are equally relevant to the performance of the neural predictor.

In the neural networks for system identification context, there has been a considerable interest in on-line training algorithms so as to cope with unmodelled dynamics and variant systems. In this perspective, several gradient-descent based algorithms have been proposed such as the basic Real-Time Recurrent Learning (RTRL) algorithm [8], further improved with respect to the learning speed and convergence [9] and the Backpropagation Through Time (BPTT) [10] with a history cut-off a finite number of time-steps: the truncated BPTT [11]. Because these methods are based on a gradient search direction, they are very often slow in reaching a satisfactory solution and additionally the training performance is quite sensitive to the learning rate choice.

In real-time learning, these issues have been to some extent addressed by regarding the recurrent neural networks training as a non-linear parameter estimation, being the extended Kalman filter (EKF) [12], [13],[14], in its many forms, one of the most widely used methods. Nevertheless, it is well known that as a result of a first-order non-linear system's approximation large errors in the true posterior mean and covariance of the transformed Gaussian variable can be introduced, which may lead to the EKF instability. This is especially evident when the model is highly non-linear and therefore the effects of higher order terms of the Taylor series expansion cannot be neglected.

Recently, a new generalization of the Kalman filter to non-linear systems on the basis of the unscented transformation (UT) has been proposed (UKF) [15], for which is claimed to be accurate to the third-order for Gaussian distributions and any kind of non-linearities [16]. Unlike the EKF, which makes use of non-linear system first-order approximations, the UKF approximates the random variable distribution by generating a discrete distribution comprising the minimum number of points that preserves the same first and second order moments.

Motivated by the identification problem of non-linear time variant systems, this paper explores the application of the UKF to the on-line non-linear state-space neural network parameters estimation. In section 2 it is presented the class of recurrent neural networks considered for modelling purpose. Section 3 describes the unscented transformation and the new filter algorithm. The fourth section is devoted to the application of the proposed on-line identification methodology to a laboratory heating system and experimental results presented and discussed. Finally, section 5 provides some concluding remarks.

2. State Space Neural Networks

In this work the black-box model is derived by means of a hybrid recurrent neural network comprising 3 layers, as depicted in Fig. 1. The input and output layers incorporate as much neurons as the number of inputs and outputs of the system to be modelled, whereas the number of neurons in the hidden layer should be the most appropriate to achieve a good approximation. In view of selecting the optimal number of hidden neurons, one should mention, without going into details, that there is always an inevitable trade-off between generalization performance and training error. In fact, a neural network larger than might be expected can result in a considerable specialisation (overfitting), which leads to poor generalization capabilities. On the other hand, if the number of neurons in the hidden layer is too small it may not be feasible to train appropriately the neural network to represent satisfactory a given data set.

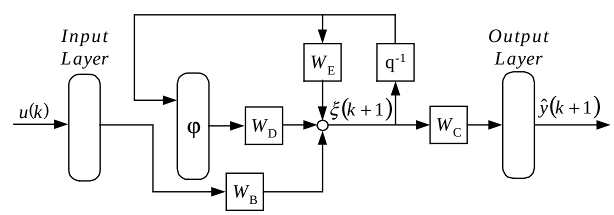


Fig. 1. State-space neural network block diagram.

In this neural network topology $\xi \in \Re^{Nn}$ denotes the neural state-space vector, $\hat{y} \in \Re^{No}$ is the neural output, $u \in \Re^{Ni}$ is the neural external input; Nn, No and Ni are, respectively, the number of neurons in the hidden layer, output layer and input layer; φ is a non-linear activation function; q^{-1} denotes the backward shift operator. Additionally, the synaptic weights between neurons: W_B, W_C, W_D and W_E are real-valued matrices having appropriate dimensions.

The state-space model resulting from the dynamic neural network architecture may be described by the following non-linear difference equations:

$$\begin{aligned} \xi(k+1) &= W_D \tanh(\xi(k)) + W_E\, \xi(k) + W_B\, u(k);\ \xi(0) = \xi_0 \\ \hat{y}(k) &= W_C\, \xi(k) \end{aligned} \tag{1}$$

assuming a hyperbolic tangent as the activation function.

3. The Uncertainty Transformation and The Unscented Kalman Filter

The unscented transformation enables the computation of the statistics of a random variable propagated through a non-linear mapping.

Consider then a Nw-dimensional real-valued random variable w with mean $\overline{w}$ and covariance matrix P_{ww} and suppose that it is required to predict the mean and covariance of $y \in \Re^q$ given as:

$$y = h(w) \tag{2}$$

with $h : \Re^{Nw} \to \Re^q$.

Firstly, a set of $2Nw+1$ pairs of weights and translated sigma points (Γ_i, ω_i) is formed according to (3), in such a way that the mean and the covariance of w are preserved.

$$\begin{cases} \left(\kappa(N+\kappa)^{-1}, \overline{w}\right), \; i = 0 \\ \left(0.5(N+\kappa)^{-1}, \overline{w} + \sqrt{(N+\kappa)(P_{ww})_i}\right), \; i = 1, \ldots, Nw \\ \left(0.5(N+\kappa)^{-1}, \overline{w} - \sqrt{(N+\kappa)(P_{ww})_i}\right), \; i = 1, \ldots, Nw \end{cases} \tag{3}$$

with $(\mathrm{P}_{ww})_i$ the i^{th} column or row of covariance matrix P_{ww} and κ a scaling parameter. These sigma points are then subsequently propagated through the non-linear mapping,

$$Y_i = h(\omega_i) \tag{4}$$

and the corresponding mean and covariance computed as follows:

$$\begin{aligned} \overline{y} &= \sum_{i=0}^{2Nw} \Gamma_i Y_i \\ \mathrm{P}_{yy} &= \sum_{i=0}^{2Nw} \Gamma_i \left(Y_i - \overline{y}\right)\left(Y_i - \overline{y}\right)^T \end{aligned} \tag{5}$$

The unscented Kalman filter consists in the application of UT to the recursive estimation of non-linear discrete-time dynamic system parameters. In the present work, the non-linear system is assumed to be described as:

$$\begin{aligned} x(k+1) &= f\left[x(k), u(k), w(k), k\right] + \upsilon(k) \\ z(k) &= h\left[x(k), w(k), k\right] + \eta(k) \end{aligned} \tag{6}$$

where $z \in \Re^p$ denotes the observation vector; $\upsilon \in \Re^n$ is the process noise with positive semi-definite covariance matrix $Q \in \Re^{n \times n}$ and $\eta \in \Re^p$ the measurement noise with positive semi-definite covariance matrix $R \in \Re^{n \times n}$.

Like all Kalman filter based algorithms, in the UKF approach the estimates are computed in two stages. Regarding the neural network weight estimation, in the time update stage one step-ahead prediction of the model parameters and the covariance of the weight-estimate error P_{ww} are evaluated, whereas in the measurement update these estimates are subsequently refined on the basis of the most recent observations. The overall filter equations are given by:

Time update stage:

$$\begin{aligned} &\Omega(k \mid k-1) = \Omega(k-1 \mid k-1) \\ &\mathrm{P}_{ww}(k \mid k-1) = \mu^{-1}\, \mathrm{P}_{ww}(k-1 \mid k-1) \\ &Z^w(k \mid k-1) = h\left(\hat{x}(k-1 \mid k-1), \Omega(k \mid k-1), k\right) \\ &\hat{z}_w(k \mid k-1) = \sum_{i=0}^{2Nw} \Gamma_i^w Z_i^w(k \mid k-1) \end{aligned} \tag{7}$$

where Ω denotes the sigma point matrix and Γ the associated weight vector.

Measurement update stage:

$$\begin{aligned} P_{\nu\nu}^w(k \mid k-1) = \sum_{i=0}^{2Nw} \Big\{ \Gamma_i^w \left[Z_i^w(k \mid k-1) - \hat{z}_w(k \mid k-1)\right] \\ \left[Z_i^w(k \mid k-1) - \hat{z}_w(k \mid k-1)\right]^T \Big\} + R \end{aligned} \tag{8}$$

$$\begin{aligned} P_{wz_w}(k \mid k-1) = \sum_{i=0}^{2Nw} \Big\{ \Gamma_i^w \left[\Omega_i(k \mid k-1) - \hat{w}(k \mid k-1)\right] \\ \left[Z_i^w(k \mid k-1) - \hat{z}(k \mid k-1)\right]^T \Big\} \end{aligned}$$

$$K^w(k) = \mathrm{P}_{wz_w}(k \mid k-1)\left(P_{\nu\nu}^w(k \mid k-1)\right)^{-1}$$

$$\hat{w}(k \mid k) = \hat{w}(k \mid k-1) + K^{w}(k)\left[z(k) - \hat{z}(k \mid k-1)\right]$$

$$P_{ww}(k \mid k) = P_{ww}(k \mid k-1) - K^{w}(k) P_{vv}^{w}(k \mid k-1) K^{w}(k)^{T}$$

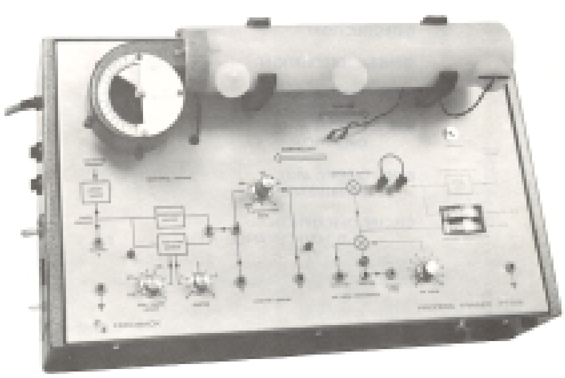
with K the Kalman gain.

4. Application

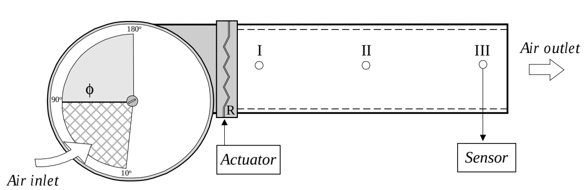
4.1. The Experimental Set-up

The process used for assessing the proposed on-line identification scheme is the laboratory heating system illustrated in **Error! Reference source not found.**.a and 2.b.

Air drawn from the local atmosphere is forced to circulate by means of a centrifugal fan through a finite length of duct and driven to the local atmosphere again. The air flux is heated in a heater grid located just after the fan outlet and its temperature measured at one of the three points available along the tube.



a) Picture.



b) Schematics.

Fig. 2. The heating system.

This is a non-linear process with a pure time delay that depends on the position of the temperature sensor and the air flow rate, which is a function of the damper position ϕ. The input to the system is a voltage on the heating device consisting of a mesh of resistor wires and the system's output is the outlet air flow temperature.

4.2. Experiments and Results

Several experiments were designed and conducted on the laboratory heating system in order to assess the feasibility, in practice, of neural networks on-line identification by means of an UKF and to emphasize how crucial the real-time parameters updating is in time variant system identification.

The neural network used for on-line identification purpose has the form of that depicted in Fig. 1. The input and output layers comprise a single neuron while in the hidden layer 3 neurons were considered.

In the first experiment, the neural network weights were randomly initialised, which reflects no previous knowledge about the system's dynamics, and the training carried out in a recursive fashion by assuming a sampling time of 0.15 second. Regarding the laboratory heating system configuration, the air damper position was set as 30° and the temperature sensor located at position III (279 *mm*). As can be inferred from the results plotted in Fig. 3, the system's dynamics were fairly well captured, despite a slight slower transient response of the model output.

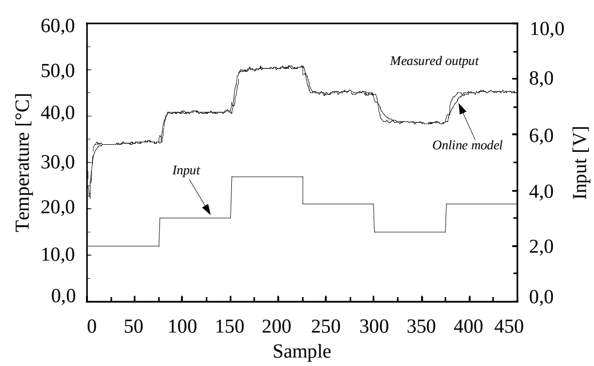


Fig. 3. Online identification: random initialisation.

Next, to assess the merits of UKF online parameters adaptation in a post-training framework, the neural network was first trained offline by considering a set of training data, previously collected from the heating system, and using the Levenberg-Marquardt algorithm to the minimisation of the prediction error. The model parameters were then subsequently used for further adaptation within the UKF online training.

As can be shown from Fig. 4, though the neural predictor time response based on batch training is somehow acceptable for lower operating points, it is quite clear that by adjusting in real-time the neural network weights a better description of the system's dynamics can be achieved regardless the operating range. This feature is particularly important when it is required the neural predictor to reflect changes in the system's dynamics due either to a structural modification or as a result of time-variant parameters of the plant.

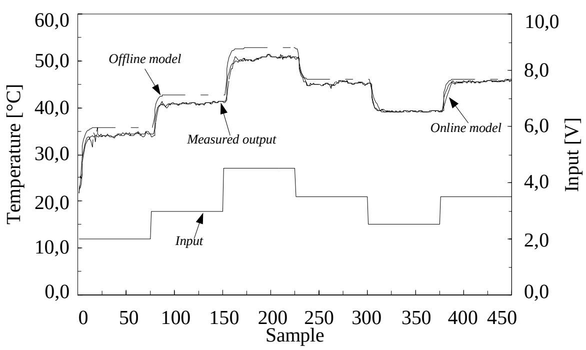


Fig. 4. Batch training plus real time weights adaptation.

To illustrate the importance of real-time training when the system is subject to a structural variation, an experiment was carried out on the laboratory process where the damper position initially located at 30° was manually changed to 40° at around the 68th sample. As a result, the air flux is increased, which leads to a decreasing in the system time delay and reducing the outlet temperature.

From the results shown in Fig. 5 it is clear that the offline model was unable to reflect the structural modification imposed on the plant, as would be expected since the neural network was trained to replicate another mapping. Thus, in this case, the neural network provides an unacceptable prediction of the heating system time response. On the contrary, by an on-line identification approach the new system's dynamics is adequately retrieved.

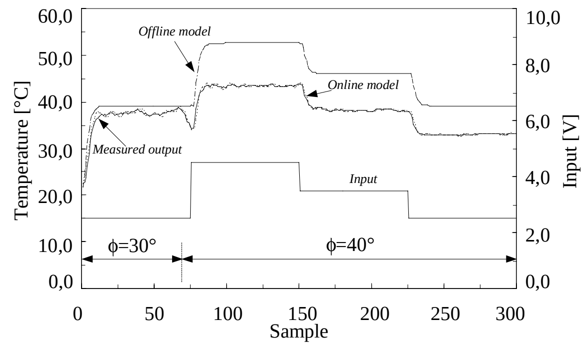


Fig. 5. Variant system: offline versus online weights adjusting.

5. Conclusions

In this paper the application of a state-space neural network as a means for modelling dynamic systems together with an unscented Kalman filter to adjust online the network's weights has been presented. Given the simplicity of this network topology along with the fast training and reliability provided by the online learning algorithm it is a viable and encouraging alternative for real-time system identification, as proved by the set of experiments conducted on the laboratory plant. In this context, given the adaptive features revealed by the state-space neural network, as well as their ability for modelling non-linear time-variant plants, the presented methodology provides a surplus value to the control field and particularly to those strategies using an explicit model of the plant to be controlled such as the model-based predictive control or the output regulation theory.

Acknowledgements

This work is partially supported by POSI - Programa Operacional da Sociedade de Informação of Portuguese Fundação para a Ciência e Tecnologia and European Union FEDER and PRAXIS XXI ALCINE Project.

References

[1] L. Ljung, Development of System Identification, *Proc. 13th World Congress of IFAC*, Vol. G, pp. 141-146, San Francisco, California, 1996.

[2] K. Narendra and K. Parthasarathy, Identification and Control of Dynamical Systems Using Neural Networks, *IEEE Trans. on Neural Networks*, vol. 1, no. 1, pp. 4-27, 1990.

[3] J. Suykens, B. De Moor, J. Vandewalle, Nonlinear System Identification Using Neural State Space Models Applicable to Robust Control Design, *Int. Journal of Control*, vol. 62, no. 1, pp. 129-152, 1995.

[4] C. Bailer-Jones, D. MacKay and P. Withers, A Recurrent Neural Network for Modelling Dynamical Systems, *Network: Computation in Neural Systems*, vol. 9, pp. 531-547, 1998.

[5] C. Gan and K. Danai, Model-Based Recurrent Neural Network for Modeling Nonlinear Dynamic System, *IEEE Int. Conf. on Control Application*, pp. 1749-1754, Hawai, 1999.

[6] I. Rivals and L. Personnaz, Black Box Modeling with State-Space Neural Networks, *Neural Adaptive Control Technology*, R. Zbikowski and K. Hunt eds., World Scientific, pp. 237-264, 1996

[7] B. Pearlmutter, Gradient Calculations for Dynamic Recurrent Neural Networks: A Survey, *IEEE Trans. on Neural Networks*, vol. 6, no. 5, 1995.

[8] R. Williams and D. Zipser, Gradient-Based Learning Algorithms for Recurrent Networks and Their Computational Complexity, In Back-propagation: Theory and Applications, Y, Chauvin and D. Rumelhart eds., Hillsdale, NJ: Erlbaum, 1995.

[9] M. Mak, K. Ku and Y. Lu, On the improvement of the real time recurrent learning algorithm for recurrent neural networks, *Neurocomputing*, vol. 24, pp. 13-36, 1999.

[10] P. Werbos, Backpropagation Through Time: What It Does and How to Do It, *Proc. of the IEEE*, vol. 78, no. 10, pp. 1550-1560, 1990.

[11] R. Williams and J. Peng, An Efficient Gradient-Based Algorithm for On-line Training of Recurrent Neural Networks Trajectories, *Neural Computation*, 2, pp. 490-501, 1990.

[12] R. Williams, Training Recurrent Networks Using the Extended Kalman Filter, *Proc. of the IJCNN'92*, vol. IV, pp. 241-246, Baltimore, MD, 1992.

[13] G. Puskorius and L. Feldkamp, Neurocontrol of Nonlinear Dynamical Systems with Kalman Filter Trained Recurrent Networks, *IEEE Trans. On Neural Networks*, vol. 5, no.2, 1994.

[14] E. Plumer, Training Neural Networks Using Sequential-Update Forms of The Extended Kalman Filter, LA-UR-95-422 C.2, Los Alamos National Laboratory, 1995.

[15] S. Julier, J. Uklmann and H. Durrant-whyte, A new Methods for Nonlinear Transformation of Means and Covariances in Filters and Estimators, *IEEE Trans. on Automatic Control*, vol. 45, no. 3, 2000.

[16] E. Wan and R. Merwe, The Unscented Kalman Filter for Nonlinear Estimation, *Proc. of IEEE Symposium on Adaptive Systems for Signal Processing, Communication and Control*, Lake Louise, Canada, 2000.