Proceedings of the 35th
Conference on Decision and Control
Kobe, Japan • December 1996

Neural Approximators for Nonlinear Sliding–Window State Observers

A. Alessandri *, M. Maggiore *, T. Parisini **, and R. Zoppoli *

* Department of Communications, Computer and System Sciences
DIST-University of Genoa, Via Opera Pia 13, 16145 Genova, Italy

** Department of Electrical, Electronic and Computer Engineering
DEEI–University of Trieste, Via Valerio 10, 34175 Trieste, Italy
E-mail: thomas@dist.unige.it

I. Some preliminary issues on deterministic nonlinear state estimation

Let us consider the discrete-time dynamic system

$$\underline{x}_{t+1} = \underline{f}(\underline{x}_t, \underline{u}_t), \quad t = 0, 1, \ldots \tag{1}$$

$$\underline{y}_t = \underline{h}(\underline{x}_t), \quad t = 0, 1, \ldots \tag{2}$$

where $\underline{x}_t \in \mathbb{R}^n$, $\underline{u}_t \in \mathbb{R}^m$, and $\underline{y}_t \in \mathbb{R}^p$ are the state, control, and measurement vectors, respectively. The initial state $\underline{x}_0$ is unknown. We assume that $\underline{x}_0 \in X$ and $\underline{u}_t \in U$, where X and U are compact sets.

Now, let us consider a *sliding-window* observer. This means that, at a given stage t and for a given temporal window of length N stages, we have to recovery $\underline{x}_{t-N}$ on the basis of the last $N+1$ measurement vectors $\underline{y}_{t-N}, \ldots, \underline{y}_t$ and the last N control vectors, $\underline{u}_{t-N}, \ldots, \underline{u}_{t-1}$. For $t = N, N+1, \ldots$, let us introduce the following systems of nonlinear equations

$$\underline{H}\left(\underline{x}_{t-N}, \underline{u}_{t-N}^{t-1}\right) \triangleq \begin{bmatrix} \underline{h}\left(\underline{x}_{t-N}\right) \\ \underline{h} \circ \underline{f}^{\underline{u}_{t-N}}(\underline{x}_{t-N}) \\ \vdots \\ \underline{h} \circ \underline{f}^{\underline{u}_{t-1}} \circ \cdots \circ \underline{f}^{\underline{u}_{t-N}}(\underline{x}_{t-N}) \end{bmatrix} = \begin{bmatrix} \underline{y}_{t-N} \\ \underline{y}_{t-N+1} \\ \vdots \\ \underline{y}_t \end{bmatrix} \tag{3}$$

where "$\circ$" denotes composition, $\underline{u}_{t-N}^{t-1} \triangleq \text{col}\{\underline{u}_{t-N}, \underline{u}_{t-N+1}, \ldots, \underline{u}_{t-1}\}$ (similarly, in the following, we let $\underline{y}_{t-N}^{t} \triangleq \text{col}\{\underline{y}_{t-N}, \underline{y}_{t-N+1}, \ldots, \underline{y}_t\}$), and, as in [1], $\underline{f}^{\underline{u}_i}(\underline{x}_i) \triangleq \underline{f}(\underline{x}_i, \underline{u}_i)$.

To state the estimation problem in a time–invariant context, we need that, besides U, also X be time–invariant. This is ensured by the following

Assumption 1. For any $\underline{x} \in X$ and for any $\underline{u} \in U$, $\underline{f}(\underline{x}, \underline{u}) \in X$.

The following observability definition can now be stated [1].

Definition. *The system (1) and (2) is uniformly $N+1$–observable with respect to X and U if there exists an integer N such that, for any $\underline{u}_{t-N}^{t-1} \in U^N$, the mapping $\underline{H}\left(\underline{x}_{t-N}, \underline{u}_{t-N}^{t-1}\right) : X \longrightarrow \mathbb{R}^{p(N+1)}$ is injective.*

In order to test the above observability property, we can use the following global univalence sufficient conditions [2].

Theorem 1. *Suppose that, for any $\underline{u}_{t-N}^{t-1} \in U^N$, the mapping $\underline{H}\left(\underline{x}_{t-N}, \underline{u}_{t-N}^{t-1}\right)$ is differentiable with respect to $\underline{x}_{t-N} \in X$ and define the Jacobian matrix $D(\underline{x}_{t-N}, \underline{u}_{t-N}^{t-1}) \triangleq \dfrac{\partial H}{\partial \underline{x}_{t-N}} \in \mathbb{R}^{p(N+1)\times n}$, $\underline{x}_{t-N} \in X$. Then, for any stage $t \geq N$, the following two cases can be addressed:*

1) If there exists an integer $N \geq n$ such that $n = p(N+1)$ and X is a rectangular set, then, for any $\underline{u}_{t-N}^{t-1} \in U^N$, $\underline{H}$ is globally univalent on X if $D(\underline{x}_{t-N}, \underline{u}_{t-N}^{t-1})$ is either a P–matrix or an N–matrix, $\forall\, \underline{x}_{t-N} \in X$, $\underline{u}_{t-N}^{t-1} \in U^N$.

2) If an integer $N \geq n$ such that $n = p(N+1)$ does not exist, and X is a convex set, then, for any $\underline{u}_{t-N}^{t-1} \in U^N$, $\underline{H}$ is globally univalent on X if there exists a matrix $A \in \mathbb{R}^{n\times p(N+1)}$ such that $C(\underline{x}_{t-N}, \underline{u}_{t-N}^{t-1}) \triangleq AD(\underline{x}_{t-N}, \underline{u}_{t-N}^{t-1})$ has the following (row) Diagonal Dominance Property: $|c_{ii}(\underline{x}_{t-N}, \underline{u}_{t-N}^{t-1})| > \sum_{j:j\neq i} |c_{ij}(\underline{x}_{t-N}, \underline{u}_{t-N}^{t-1})|$, $\forall\, \underline{x}_{t-N} \in X$, where c_{ij} denotes the i-th row, j-th column element of matrix C.

Let us now define the set $Y \triangleq \underline{h}(X)$, $Y \subset \mathbb{R}^p$. The $N+1$–observability of (1),(2) implies the possibility of solving the nonlinear system (3) uniquely for any vector in $Y^{N+1} \times U^N$ and for any stage $t \geq N$. This means the existence of the time–invariant mapping $\underline{x}_{t-N} = \underline{\gamma}(\underline{y}_{t-N}^t, \underline{u}_{t-N}^{t-1})$, which constitutes an order $N+1$ dead–beat observer for (1),(2). Clearly, in the general nonlinear case, computing $\underline{\gamma}(\underline{y}_{t-N}^t, \underline{u}_{t-N}^{t-1})$ in analytical form is a hard, almost impossible task. In [1], a Newton's algorithm to solve (3) is described. Under suitable assumptions, it is shown that this algorithm gives rise to an asymptotic observer for (1),(2).

II. State estimation on the basis of noisy measures

Let us consider the case in which an additive noise affects the measurement channel. Then, (2) becomes

$$\underline{y}_t = \underline{h}(\underline{x}_t) + \underline{\eta}_t, \quad t = 0, 1, \ldots \tag{4}$$

We assume the statistics of the random sequence $\{\underline{\eta}_t, t = 0, 1, \ldots\}$ to be unknown. However, we also assume that $\underline{\eta}_t \in E \subset \mathbb{R}^r$, where E is a known compact set. As an exact recovery of the state vector is now impossible, by following a traditional least–squares approach, for $t = N, N+1, \ldots$, we define the following sliding–window estimation error:

$$J_t = \mu \left\| \hat{\underline{x}}_{t-N,t} - \bar{\underline{x}}_{t-N} \right\|^2 + \sum_{i=t-N}^{t} \left\| \underline{y}_i - \underline{h}\left(\hat{\underline{x}}_{it}\right) \right\|^2 \tag{5}$$

where $\hat{\underline{x}}_{it}$, $i = t-N, \ldots, t$, is the estimate of $\underline{x}_i$ derived at the stage t on the basis of $\underline{y}_{t-N}^t$, $\underline{u}_{t-N}^{t-1}$, and the prediction $\bar{\underline{x}}_{t-N}$. μ is a positive scalar that expresses our belief in the ratio between the prediction error and the magnitudes of

0-7803-3590-2/96 $5.00 © 1996 IEEE

the measurement noises. $\bar{\underline{x}}_0$ is an arbitrary vector belonging to X.

Now, let us define the compact set $Y \triangleq \{\underline{y} \in \mathbb{R}^p : \underline{y} = \underline{h}(\underline{x}) + \underline{\eta}, \forall \underline{x} \in X, \forall \underline{\eta} \in E\}$. Define also the set of optimal predictions $\bar{\underline{x}}_t^\circ$ (see Problem 1 below) as $\mathcal{X}_t^p$. Then, we can state the following

Problem 1. *At any stage* $t = N, N+1, \ldots,$ *find the optimal state estimator* $\hat{\underline{x}}_{t-N,t}^\circ = \underline{\gamma}_{t-N}^\circ(\bar{\underline{x}}_{t-N}^\circ, \underline{y}_{t-N}^t, \underline{u}_{t-N}^{t-1}) : \mathcal{X}_t^p \times Y^{N+1} \times U^N \longrightarrow \mathbb{R}^n$ *that minimizes the cost (5) under the constraints*

$$\hat{\underline{x}}_{i+1,t} = \underline{f}(\hat{\underline{x}}_{it}, \underline{u}_i), \quad i = t-N, \ldots, t-1 \tag{6}$$

The minimizations are linked sequentially by the optimal predictions

$$\bar{\underline{x}}_{t-N}^\circ = \bar{\underline{x}}_{t-N,t-1}^\circ \triangleq \underline{f}(\hat{\underline{x}}_{t-N-1,t-1}^\circ, \underline{u}_{t-N-1}), \quad t = N+1, N+2, \ldots, \forall \bar{\underline{x}}_0 \in X \tag{7}$$

Solving Problem 1 enables one to generate a sequential state estimator.

III. A condition for the convergence of the state estimate

Define the set of optimal estimates $\hat{\underline{x}}_t^\circ$ as X_t° and introduce the following

Assumption 2. There exists a compact set $\tilde{X}$ such that $\tilde{X} \supseteq X \bigcup \left(\bigcup_{t=0}^{+\infty} X_t^\circ\right)$.

Let us now introduce some notations and useful quantities. Given a matrix $A = A^T \geq 0$, let us denote by $\lambda_{\min}(A)$ and $\lambda_{\max}(A)$ the minimum and maximum eigenvalues of A, respectively. Given a generic matrix B, $\|B\|_{\max} \triangleq \|B\| = \sqrt{\lambda_{\max}(B^T B)}$ and $\|B\|_{\min} \triangleq \sqrt{\lambda_{\min}(B^T B)}$. Further, let us define the set $\mathcal{X}$ as the closed convex hull of $\tilde{X}$. Let us assume $\underline{f}$ and $\underline{h}$ to be Lipschitz functions on $\mathcal{X}$ with the constants k_f and k_h, respectively. Moreover, define $\Delta \triangleq \max_{\underline{x}_{t-N} \in \mathcal{X}, \underline{u}_{t-N}^{t-1} \in U} \|D(\underline{x}_{t-N}, \underline{u}_{t-N}^{t-1})\|$, $\delta \triangleq \min_{\underline{x}_{t-N} \in \mathcal{X}, \underline{u}_{t-N}^{t-1} \in U} \|D(\underline{x}_{t-N}, \underline{u}_{t-N}^{t-1})\|_{\min}$, $\tilde{k} = \Delta / k_f$ and $\bar{k} \triangleq \sqrt{\bar{k}_1^2 + \ldots + \bar{k}_{N+1}^2}$, where $\bar{k}_i > 0$ are suitable scalars such that $\|\frac{\partial \underline{h}}{\partial \underline{x}_{t-N}}(\underline{x}'_{t-N}) - \frac{\partial \underline{h}}{\partial \underline{x}_{t-N}}(\underline{x}''_{t-N})\| \leq \bar{k}_1 \|\underline{x}'_{t-N} - \underline{x}''_{t-N}\|$, $\|\frac{\partial \underline{h} \circ \underline{f}^{\underline{u}_{t-N}}}{\partial \underline{x}_{t-N}}(\underline{x}'_{t-N}) - \frac{\partial \underline{h} \circ \underline{f}^{\underline{u}_{t-N}}}{\partial \underline{x}_{t-N}}(\underline{x}''_{t-N})\| \leq \bar{k}_2 \|\underline{x}'_{t-N} - \underline{x}''_{t-N}\|, \ldots, \|\frac{\partial \underline{h} \circ \underline{f}^{\underline{u}_{t-1}} \circ \cdots \circ \underline{f}^{\underline{u}_{t-N}}}{\partial \underline{x}_{t-N}}(\underline{x}'_{t-N}) - \frac{\partial \underline{h} \circ \underline{f}^{\underline{u}_{t-1}} \circ \cdots \circ \underline{f}^{\underline{u}_{t-N}}}{\partial \underline{x}_{t-N}}(\underline{x}''_{t-N})\| \leq \bar{k}_{N+1} \|\underline{x}'_{t-N} - \underline{x}''_{t-N}\|$, $\forall\, \underline{x}'_{t-N}, \underline{x}''_{t-N} \in \mathcal{X}; \forall\, \underline{u}_{t-N}^{t-1} \in U^N$ (the $\bar{k}_i$ do exist as we are addressing composition of Lipschitz functions). Finally, the following further assumption is needed:

Assumption 3. There exists an integer N such that, for any $\underline{u}_{t-N}^{t-1} \in U^N$, the mapping $\underline{H}\left(\underline{x}_{t-N}, \underline{u}_{t-N}^{t-1}\right) : \mathcal{X} \longrightarrow \mathbb{R}^{p(N+1)}$ is an injective immersion.

Then, we can state the following results.

Theorem 2. *Suppose that Assumptions 1,2, and 3 are verified. Denote by* $\underline{e}_{t-N} \triangleq \underline{x}_{t-N} - \hat{\underline{x}}_{t-N,t}^\circ$ *the estimation error at stage* $t-N$. *Consider the largest closed ball* $N(r_e)$ *of radius* r_e *and with the center in the origin such that* $\underline{e}_0 \in N(r_e)$, *and define the scalar* $r_\eta \triangleq \max_{\underline{\eta}_{t-N}, \ldots, \underline{\eta}_t \in E^{N+1}} \|\mathrm{col}(\underline{\eta}_{t-N}, \ldots, \underline{\eta}_t)\|$. *If there exists a choice of* μ *for which the inequalities*

$$(1-k_f)^2 \mu^3 + (3 + k_f^2 - 4k_f)\delta^2 \mu^2 + (3\delta^4 - 2k_f \delta^4 - 8\bar{k}\tilde{k}^2 k_f^3 r_\eta)\mu + \delta^6 > 0 \tag{8}$$

$$(k_f - 1)\mu < \delta^2 \tag{9}$$

are satisfied, the second-order equation

$$(\bar{k}\tilde{k}k_f^2\mu^2)\xi^2 + [k_f\mu^2(\mu+\delta^2) - \mu(\mu+\delta^2)^2 + 2\bar{k}\tilde{k}^2 k_f^2 \mu r_\eta]\xi + \bar{k}\tilde{k}^3 k_f^2 r_\eta^2 + \tilde{k}k_f\mu(\mu+\delta^2) = 0$$

has the two real positive roots ξ^- *and* ξ^+, *with* $\xi^- < \xi^+$.

Then, if the choice of μ *yields also the fulfilment of the inequality*

$$\mu^2 + 2\left(\delta^2 - \bar{k}\,\tilde{k}\,k_f^2\,\xi^+\right)\mu + \left(\delta^4 - 2\bar{k}\,\tilde{k}^2\,k_f^2\,r_\eta\right) > 0 \tag{10}$$

we have

$$\lim_{t \to +\infty} \|\underline{e}_t\| \leq \xi^-, \quad \forall\, r_e < \xi^+ \tag{11}$$

It is worth noting that Assumption 2 has been introduced such that the constants k_f, k_h, δ, and Δ are well-defined on the compact set $\mathcal{X}$, but it does not influence the convergence of the estimation error. In other terms, if the inequalities introduced in Theorem 2 are satisfied so that the bound (11) holds true, it is easy to show that Assumption 2 is verified.

As the statement of Problem 1 does not impose any particular way of computing $\hat{\underline{x}}_{t-N}^\circ$, we have two possibilities:

1) *On-line computation.* Problem 1 can be regarded as a nonlinear programming one. The main advantage of this approach is that many well-established nonlinear programming techniques are available to solve this problem.

2) *Off-line computation.* This approach implies that the function $\underline{\gamma}_{t-N}^\circ(\bar{\underline{x}}_{t-N}^\circ, \underline{y}_{t-N}^t, \underline{u}_{t-N}^{t-1})$ has to be computed "a priori" and stored in the observer's memory. This enables the optimal estimates to be determined on line "instantaneously". Clearly, the off-line computation has advantages and disadvantages that are opposite to the ones of the on-line approach. No on-line computational effort is requested from the observer, but a large amount of computer memory may be required to store the estimation law.

IV. The neural approximation for the nonlinear observer and a simulation example

To state the problem in a time-invariant context, we need a further assumption (analogous to Assumption 2).

Assumption 4. There exists a set $\mathcal{X}^p$ such that $\mathcal{X}^p \supseteq \left(\bigcup_{t=0}^{+\infty} \mathcal{X}_t^p\right)$.

If Assumption 4 is verified, in virtue of the time-invariance of (1),(2), and (5), the index $t-N$ can be dropped from the function $\underline{\gamma}_{t-N}^\circ$ and the stationary estimation law $\underline{\gamma}^\circ(\bar{\underline{x}}_{t-N}^\circ, \underline{y}_{t-N}^t, \underline{u}_{t-N}^{t-1})$ can be addressed. Then, without loss of generality, we consider the stage $t = N$, hence we state Problem 1 in the following equivalent way:

Problem 1'. *Find the optimal estimation law* $\underline{\hat{x}}_0^\circ = \underline{\gamma}^\circ(\underline{\bar{x}}_0^\circ, \underline{y}_0^N, \underline{u}_0^{N-1})$ *that minimizes the cost (5) under the constraints (6) for any* $\underline{\bar{x}}_0^\circ \in \mathcal{X}^p, \underline{y}_0^N \in Y^{N+1}, \underline{u}_0^{N-1} \in U^N$.

As Problem 1' cannot in general be solved analytically, we propose to approximate $\underline{\gamma}^\circ(\underline{\bar{x}}_0^\circ, \underline{y}_0^N, \underline{u}_0^{N-1})$ by a function $\hat{\underline{\gamma}}(\underline{\bar{x}}_0^\circ, \underline{y}_0^N, \underline{u}_0^{N-1}, \underline{w})$, to which we assign a given structure. $\underline{w}$ is a vector of parameters to be optimized. Among various possible approximating functions, we choose multilayer feedforward neural networks (in this case, $\underline{w}$ is the vector of the synaptic weights).

Theorem 2 can be extended [3] to the case of approximate solutions of Problem 1 (i.e., to Problem 1'). Shortly put, a positive scalar $\bar{\varepsilon}$, denoting the maximum approximation error, can be computed for which a result analogous to that of Theorem 2 holds true. Therefore, we have to specify the magnitudes of the errors generated by the law $\underline{\hat{x}}_0 = \hat{\underline{\gamma}}(\underline{\bar{x}}_0^\circ, \underline{y}_0^N, \underline{u}_0^{N-1}, \underline{w})$. To this end, we assume that the approximating neural function contains only one hidden layer composed of ν neural units, and that the output layer is composed of linear activation units. We denote such a function by $\hat{\underline{\gamma}}^{(\nu)}(\underline{\bar{x}}_0^\circ, \underline{y}_0^N, \underline{u}_0^{N-1}, \underline{w})$. Then, we can state the following

Theorem 3. *Assume that the optimal estimation function* $\underline{\gamma}^\circ(\underline{\bar{x}}_0^\circ, \underline{y}_0^N, \underline{u}_0^{N-1})$ *is unique and that it is a* $\mathcal{C}[\mathcal{X}^p \times Y^{N+1} \times U^N, \mathbb{R}^n]$ *function. Then, for any* $\varepsilon \in \mathbb{R}, \varepsilon > 0$, *there exist an integer* ν *and a weight vector* $\underline{w}$ *such that*

$$\left\| \underline{\gamma}^\circ(\underline{\bar{x}}_0^\circ, \underline{y}_0^N, \underline{u}_0^{N-1}) - \hat{\underline{\gamma}}^{(\nu)}(\underline{\bar{x}}_0^\circ, \underline{y}_0^N, \underline{u}_0^{N-1}, \underline{w}) \right\| < \varepsilon ,$$
$$\forall\, \underline{\bar{x}}_0^\circ \in \mathcal{X}^p,\ \underline{y}_0^N \in Y^{N+1},\ \underline{u}_0^{N-1} \in U^N \qquad (12)$$

The above theorem has been derived from a well-known property, according to which continuous functions can be approximated, to any degree of accuracy, on a given compact set by feedforward neural networks based on sigmoidal functions, provided that the number ν of neural units is sufficiently large. In order to guarantee the aforementioned uniform bound $\bar{\varepsilon}$ to the approximation error, the following min-max problem is stated.

Problem 2. *Find the number* ν^* *of neural units such that*

$$\min_{\underline{w}} \max_{\underline{\bar{x}}_0 \in \mathcal{X}^p, \underline{y}_0^N \in Y^{N+1}, \underline{u}_0^{N-1} \in U^N} \| \underline{\gamma}^\circ(\underline{\bar{x}}_0, \underline{y}_0^N, \underline{u}_0^{N-1}) - \hat{\underline{\gamma}}^{(\nu)}(\underline{\bar{x}}_0, \underline{y}_0^N, \underline{u}_0^{N-1}, \underline{w}) \| \leq \bar{\varepsilon} \qquad (13)$$

As to the number ν^* of neural units, rather a naive trial-and-error procedure for determining them is the following: increase ν until the term on the left-hand side of (13) is less than or equal to $\bar{\varepsilon}$.

To show the effectiveness of the proposed neural approach to the nonlinear state estimation problem, let us consider a *Target Motion Analysis* (TMA) problem, in particular, the class of nonlinear passive tracking problems known as *Bearings Only Measurements Problems* (BOMPs). A mobile target Q is moving at velocity $\underline{v}_Q$, while an observer P, moving at velocity $\underline{v}_P$, is trying to track it by using only noisy measurements of the line-of-sight angle β. Let us denote by $\underline{r}_t$ the relative position of Q with respect to P and by $\underline{v}_t$ the corresponding relative velocity. The system can be modelled by a linear equation $\underline{x}_{t+1} = A\underline{x}_t + B\underline{u}_t + \underline{\xi}_t$, $t = 0, 1, \ldots$, where $\underline{x}_t \triangleq \mathrm{col}\,(r_{x_t}, v_{x_t}, r_{y_t}, v_{y_t})$, $\underline{u}_t \triangleq \mathrm{col}(u_{x_t}, u_{y_t})$ is the acceleration vector, and matrices A and B can be easily computed on the basis of the sampling period $T = 0.1$s. The nonlinear observation channel is given by $\beta_t = \arctan(r_{y_t}/r_{x_t}) + \eta_t$, $t = 0, 1, \ldots$. Moreover $\underline{x}_0 \sim N\left((2.2, 2, 2.2, -2)^T, \Sigma_{x_0}\right)$, $\eta_t \sim N\left(0, \sigma_\eta^2\right)$, with $\Sigma_{x_0} = \mathrm{diag}\,(0.08, 0.0008, 0.08, 0.0008)$ and $\sigma_\eta^2 = 0.3 \cdot 10^{-4}$. It is assumed that the target is travelling at a constant speed of $0.8\,\mathrm{m\,s^{-1}}$ in the 45° direction, while the observer is traversing a multileg maneuver consisting of known periodic changes in the components of $\underline{u}_t$.

The RMS estimation errors on r_{x_t} of the neural observer (derived by the mini-max technique) and of the EKF are compared in Fig. 1. Similar behaviors can be shown concerning the other state variables.

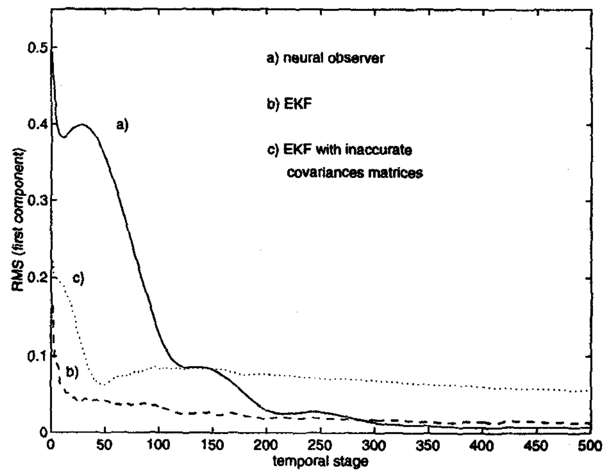

Fig. 1. Behaviors of the estimates of r_{x_t} generated by the neural observer and by the EKF.

It can be seen that the neural observer, after a certain number of temporal stages, behaves better than the EKF. This can be easily explained by considering that the EKF has a complete knowledge on the statistics of the random variables. Instead, the initial value $\underline{\bar{x}}_0$ is chosen arbitrarily, possibly quite far from the corresponding true value. Moreover, the EKF turns out to be quite sensitive to the statistics of the random variables as is shown in diagram c) where incorrect initializations of the covariances matrices are considered.

Of course, more complete theoretical and experimental results are needed to establish under what conditions (i.e., strong nonlinearities and large noises) the neural observer performs better than the EKF.

References

[1] P. E. Moraal and J. W. Grizzle, "Observer design for nonlinear systems with discrete-time measurements," *IEEE Trans. on Automatic Control*, Vol. 40, No. 3, pp. 395-404, 1995.

[2] J. M. Fitts, "On the observability of nonlinear systems with applications to nonlinear regression analysis," *Information Sciences*, Vol. 4, pp. 129-156, 1972.

[3] A. Alessandri, M. Maggiore, T. Parisini, and R. Zoppoli, "A condition for the error convergence of an asymptotic nonlinear state estimator," DIST Tech. Report No. 96/1, 1996.