Proceedings of the American Control Conference
Albuquerque, New Mexico June 1997
0-7803-3832-4/97/$10.00 © 1997 AACC

On convergence of neural approximate nonlinear state estimators[1]

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

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

** IAN–Institute for Naval Automation
CNR–National Research Council, Via De Marini 6, 16149 Genova, Italy

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

Abstract

The problem of designing a state observer for nonlinear systems has been faced in several works in the past decades and only recently researches focused on the discrete–time ones. In the paper, the case of a noisy measurement channel is addressed. By generalizing the classical least-squares method we compute the estimation law off line by solving a functional optimization problem. Convergence results of the estimation error are stated and the approximate solution of the above problem is addressed by means of a feedforward neural network. A min–max technique is proposed to determine the weight coefficients of the "neural" observer so as to estimate the system state to any given degree of accuracy, thus guaranteeing the boundedness of the estimation error.

1. Some preliminary issues on deterministic nonlinear state estimation

Let us consider the discrete-time dynamic system

$$x_{t+1} = f(x_t, u_t), \quad t = 0, 1, \ldots \tag{1}$$

$$y_t = h(x_t), \quad t = 0, 1, \ldots \tag{2}$$

where $x_t \in \mathbb{R}^n$, $u_t \in \mathbb{R}^m$, and $y_t \in \mathbb{R}^p$ are the state, control, and measurement vectors, respectively. The initial state x_0 is unknown. We assume that $x_0 \in X$ and $u_t \in U$, 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 x_{t-N} on the basis of the last $N+1$ measurement vectors $y_{t-N}, \ldots, y_t$ and the last N control vectors, $u_{t-N}, \ldots, u_{t-1}$. For $t = N, N+1, \ldots$, let us introduce the following systems of nonlinear equations

$$F\left(x_{t-N}, u_{t-N}^{t-1}\right) \triangleq \begin{bmatrix} h\left(x_{t-N}\right) \\ h \circ f^{u_{t-N}}(x_{t-N}) \\ \vdots \\ h \circ f^{u_{t-1}} \circ \cdots \circ f^{u_{t-N}}(x_{t-N}) \end{bmatrix} = \begin{bmatrix} y_{t-N} \\ y_{t-N+1} \\ \vdots \\ y_t \end{bmatrix} \tag{3}$$

where "$\circ$" denotes composition, $u_{t-N}^{t-1} \triangleq \text{col}(u_{t-N}, \ldots, u_{t-1})$ (similarly, in the following, we let $y_{t-N}^{t} \triangleq \text{col}\left(y_{t-N}, \ldots, y_t\right)$), and, as in [1], $f^{u_i}(x_i) \triangleq f\left(x_i, u_i\right)$.

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 $x \in X$ and for any $u \in U$, $f(x, u) \in X$.

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 $u_{t-N}^{t-1} \in U^N$, the mapping $F\left(x_{t-N}, u_{t-N}^{t-1}\right) : X \longrightarrow \mathbb{R}^{p(N+1)}$ 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 $u_{t-N}^{t-1} \in U^N$, the mapping $F\left(x_{t-N}, u_{t-N}^{t-1}\right)$ is differentiable with respect to $x_{t-N} \in X$ and define the Jacobian ma-*

[1] All correspondence should be mailed to Dr. Thomas Parisini. This work was partly supported by the Italian Ministry for the University and Research (MURST) and by the Italian Space Agency (ASI).

trix $D(x_{t-N}, u_{t-N}^{t-1}) \triangleq \dfrac{\partial F}{\partial x_{t-N}} \in \mathbb{R}^{p(N+1)\times n}$, $x_{t-N} \in X$. 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 $u_{t-N}^{t-1} \in U^N$, F is globally univalent on X if $D(x_{t-N}, u_{t-N}^{t-1})$ is either a P-matrix or an N-matrix, $\forall x_{t-N} \in X$, $u_{t-N}^{t-1} \in U^N$.

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 $u_{t-N}^{t-1} \in U^N$, F is globally univalent on X if there exists a matrix $A \in \mathbb{R}^{n\times p(N+1)}$ such that $C(x_{t-N}, u_{t-N}^{t-1}) \triangleq AD(x_{t-N}, u_{t-N}^{t-1})$ has the following (row) Diagonal Dominance Property: $|c_{ii}(x_{t-N}, u_{t-N}^{t-1})| > \sum_{j:j\neq i} |c_{ij}(x_{t-N}, u_{t-N}^{t-1})|$, $\forall x_{t-N} \in X$, where c_{ij} denotes the i-th row, j-th column element of matrix C.

Let us now define the set $Y \triangleq h(X)$, $Y \subset \mathbb{R}^p$. 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 $x_{t-N} = \gamma(y_{t-N}^t, u_{t-N}^{t-1})$, which constitutes an order $N+1$ dead-beat observer for (1),(2). Clearly, in the general nonlinear case, computing $\gamma(y_{t-N}^t, u_{t-N}^{t-1})$ 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).

2. 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

$$y_t = h(x_t) + \eta_t, \quad t = 0, 1, \ldots \tag{4}$$

We assume the statistics of the random sequence $\{\eta_t, t = 0, 1, \ldots\}$ to be unknown. However, we also assume that $\eta_t \in H \subset \mathbb{R}^r$, where H is a known compact set. Let us define the information vector

$$I_t^N \triangleq \text{col}\,(\bar{x}_{t-N}^\circ, y_{t-N}, \ldots, y_t, u_{t-N}, \ldots, u_{t-1}), \quad t = N, N+1, \ldots$$

where $\bar{x}_{t-N}^\circ \triangleq \hat{x}_{t-N,t-1}^\circ$ denotes the optimal prediction (see Problem 1 below). Of course, $\bar{x}_0^\circ = \hat{x}_{0,N-1} = \bar{x}_0$ denotes an *a-priori* prediction.

Let us now define as X_0 the set from which the random vector x_0 takes its value. We also label as X_t, Y_t, and $\mathcal{I}_t^N$ the sets from which the vectors x_t, y_t and I_t^N take their values, respectively. Clearly, we have

$$X_t \triangleq \begin{cases} X_0, & \text{for } t = 0 \\ f(X_{t-1}, U), & \text{for } t \geq 1 \end{cases} \tag{5}$$

$$Y_t \triangleq h(X_t, H), \text{ for } t \geq 0 \tag{6}$$

$$\mathcal{I}_t^N \triangleq \bar{X}_{t-N}^\circ \times Y_{t-N} \times \cdots \times Y_t \times (U)^{N-1} \tag{7}$$

where $\bar{X}_{t-N}^\circ$ denotes the sets the optimal predictions $\bar{x}_{t-N}^\circ$ belong to, $\forall t \geq N$.

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:

$$J_t = \mu \| \hat{x}_{t-N,t} - \bar{x}_{t-N} \|^2 + \sum_{i=t-N}^{t} \| y_i - h(\hat{x}_{it}) \|^2 \tag{8}$$

μ is a positive scalar that expresses our beliefs in the ratio between the prediction error and the measurement errors, i.e., the errors in the observation model and noises (such beliefs could be expressed more thoroughly by suitable weight matrices without additional conceptual difficulties in the reasoning reported later on). Then, we can state the following

Problem 1. *At any stage $t = N, N+1, \ldots$, find the optimal estimation function $\hat{x}_{t-N,t}^\circ = a_{t-N,t}^\circ(I_t^N)$, that minimizes the cost (8) under the constraints*

$$\hat{x}_{i+1,t} = f(\hat{x}_{it}, u_i), \quad i = t-N, \ldots, t-1 \tag{9}$$

The minimizations are linked sequentially by the optimal predictions

$$\bar{x}_{t-N}^\circ = f(\hat{x}_{t-N-1,t-1}^\circ, u_{t-N-1}), \quad t = N+1, N+2, \ldots; \bar{x}_0 \in X_0$$

□

In the next section, the estimator's convergence issue will be addressed.

3. A condition for the convergence of the state estimator that solves Problem 1

We need now the following assumptions:

A1) X_0 (see (5)), H, and U are compact sets.

A2) Problem 1 has a unique global solution, i.e., for any $t = N, N+1, \ldots$ and any $I_t^N \in \mathcal{I}_t^N$ (see (7)), there is a unique optimal estimate $\hat{x}_{t-N,t}^\circ = a_{t-N,t}^\circ(I_t^N)$.

Let us now define the sets of the optimal estimates $\hat{x}_{t-N,t}^\circ$ as $X_{t-N,t}^\circ \triangleq a_{t-N,t}^\circ(\mathcal{I}_t^N)$, $t = N, N+1, \ldots$. We need the following other assumptions:

A3) For any $x \in X_0$ and for any $u \in U$, $f(x, u) \in X_0$ (i.e., X_0 is a controlled-invariant set with respect to U).

A4) There exists a compact set X such that $X_0 \bigcup (\bigcup_{t=0}^{+\infty} X_{t-N,t}^\circ) \subseteq X$.

Finally, define $\mathcal{X}$ as the closed convex hull of X and introduce the further assumptions:

A5)The functions f and h are $\mathcal{C}^2$ functions on $\mathcal{X}$.

A6)There exists an integer N such that, for any $u_{t-N}^{t-1} \in U^N$, the mapping $F(\cdot, u_{t-N}^{t-1}) : \mathcal{X} \longrightarrow \mathbb{R}^{p(N+1)}$ is such that $\operatorname{rank}(F) = n$, $\forall x \in \mathcal{X}$.

Now, we can define the Jacobian matrix

$$D(x_{t-N}, u_{t-N}^{t-1}) \triangleq \frac{\partial F}{\partial x_{t-N}} \in \mathbb{R}^{p(N+1)\times n},\ x_{t-N} \in \mathcal{X}$$

Let us now introduce some notations and useful quantities. Given a generic symmetric positive definite matrix A, 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 $\|B\|_{\min} \triangleq \sqrt{\lambda_{\min}(B^T B)}$. Let us denote k_f and k_h the Lipschitz constants of f and h on $\mathcal{X} \times U$ and $\mathcal{X}$, respectively. Moreover, define $\Delta \triangleq \max_{x_{t-N}\in\mathcal{X}, u_{t-N}^{t-1}\in U^N} \|D(x_{t-N}, u_{t-N}^{t-1})\|$, $\delta \triangleq \min_{x_{t-N}\in\mathcal{X}, u_{t-N}^{t-1}\in U^N} \|D(x_{t-N}, u_{t-N}^{t-1})\|_{\min}$, $\tilde{k} = \Delta/k_f$ 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 $\|\frac{\partial h}{\partial x_{t-N}}(x'_{t-N}) - \frac{\partial h}{\partial x_{t-N}}(x''_{t-N})\| \leq \bar{k}_1 \| x'_{t-N} - x''_{t-N}\|$, $\|\frac{\partial h \circ f^{u_{t-N}}}{\partial x_{t-N}}(x'_{t-N}) - \frac{\partial h \circ f^{u_{t-N}}}{\partial x_{t-N}}(x''_{t-N})\| \leq \bar{k}_2 \|x'_{t-N} - x''_{t-N}\|, \ldots, \|\frac{\partial h \circ f^{u_{t-1}} \circ \cdots \circ f^{u_{t-N}}}{\partial x_{t-N}}(x'_{t-N}) - \frac{\partial h \circ f^{u_{t-1}} \circ \cdots \circ f^{u_{t-N}}}{\partial x_{t-N}}(x''_{t-N})\| \leq \bar{k}_{N+1}\|x'_{t-N} - x''_{t-N}\|$, $\forall\ x'_{t-N}, x''_{t-N} \in \mathcal{X}$; $\forall\, u_{t-N}^{t-1} \in U^N$. The $\bar{k}_i$ do exist as we are addressing compositions of $\mathcal{C}^2$ functions. Then, we can state the following result.

Theorem 2. [3] *Suppose that Assumptions A1) to A6) are verified. Denote by $e_{t-N} \triangleq x_{t-N} - \hat{x}^{\circ}_{t-N,t}$ the estimation errors at stages $t-N, t-N+1, \ldots$. Consider the largest closed ball $N(r_e)$ with radius r_e and center in the origin such that $e_0 \in N(r_e)$, and define the scalar $r_\eta \triangleq \max_{\eta_{t-N},\ldots,\eta_t \in H^{N+1}} \|\operatorname{col}(\eta_{t-N}, \ldots, \eta_t)\|$. If there exists a choice of μ, for which the inequalities*

$$(1-k_f)^2 \mu^3 + (3 + k_f^2 - 4k_f)\delta^2 \mu^2 + (3\delta^4 - 2k_f\delta^4 - 8k_f^3 \bar{k}\tilde{k}^2 r_\eta)\mu + \delta^6 > 0 \qquad (10)$$

$$(k_f - 1)\,\mu < \delta^2 \qquad (11)$$

are satisfied, the second-order equation

$$2k_f^3\bar{k}\tilde{k}\mu^2\xi^2 + [k_f\mu(\mu+\delta^2)^2 - (\mu+\delta^2)^3 + 4k_f^3\bar{k}\tilde{k}^2\mu r_\eta]\xi + 2k_f^3\bar{k}\tilde{k}^3 r_\eta^2 + k_f\tilde{k}(\mu+\delta^2)^2 r_\eta = 0 \qquad (12)$$

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

$$\mu^2 + 2\left(\delta^2 - k_f^2\bar{k}\tilde{k}\xi^+\right)\mu + \left(\delta^4 - 2k_f^2\bar{k}\tilde{k}^2 r_\eta\right) \geq 0 \qquad (13)$$

we have

$$\lim_{t\to+\infty} \|e_t\| \leq \xi^-, \quad \forall\, r_e < \xi^+ \qquad (14)$$

□

4. An approximate solution of the estimation Problem 1" by feedforward neural networks

We propose to assign the estimation functions $\hat{a}_{t-N,t}(I_t^N)$ the structure of a feedforward neural network. Let $\gamma(x) : \mathbb{R}^{n_1} \to \mathbb{R}^{n_2}$ be a generic continuous function that we want to approximate. Let us 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. This means that such a function takes on the form

$$\gamma^{(\nu)}(x,w) = \operatorname{col}\left[\gamma_j^{(\nu)}(x,w),\ j = 1, \ldots, n_2\right] \qquad (15)$$

where

$$\gamma_j^{(\nu)}(x,w) = \sum_{p=1}^{\nu} \left\{ w_{pj}(2)g\left[w_p^T x + w_{0p}(1)\right] + w_{0j}(2) \right\} \qquad (16)$$

where g is a sigmoidal activation function and $w_p = \operatorname{col}[w_{1p}(1), \ldots, w_{n_1 p}(1)]$ (clearly, w is the vector of all the weights parameters). Then, we have the following

Property 1. *Given any function $\gamma(x) \in \mathcal{C}[\mathbb{R}^{n_1}, \mathbb{R}^{n_2}]$ and any compact set $X \subset \mathbb{R}^{n_1}$, for any $\varepsilon \in \mathbb{R}$, $\varepsilon > 0$, there exists an integer ν and a weight vector w such that*

$$\left\|\gamma(x) - \gamma^{(\nu)}(x,w)\right\| \leq \varepsilon, \quad \forall x \in X \qquad (17)$$

□

Property 1 has been derived directly from the results reported, for instance, in [4], 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.

Now, as we will have to explicit Property 1, we assume that the neural estimation functions take the form (15), that is

$$\hat{x}_{t-N,t} = \hat{a}^{\nu_{t-N}}(I_t^N, w_{t-N}), \quad t = N, N+1, \ldots \qquad (18)$$

where I_t^N is the argument x, $n_1 = \dim(I_t^N)$, $n_2 = n$, and w_{t-N} are the weight vectors to be determined. To determine w_{t-N}, an approach should be followed guaranteeing that the error, due to the substitution of the estimation function $a_{t-N,t}(I_t^N)$ with the approximate function

$\hat{a}^{\nu_{t-N}}(I_t^N, w_{t-N})$, can be made uniformly bounded by a given positive scalar $\bar{\varepsilon}$. As is consistent with intuition, we need this guarantee in order to obtain convergence results for the approximate estimates similar to the ones stated in Theorem 2 (see Theorem 3 below).

First, we need to state Problem 1 in an equivalent stationary form (see [5] for more details).

Problem 1'. *Find the optimal estimation function* $\hat{x}^\circ_{t-N,t} = a^\circ(I_t^N)$, *that minimizes the cost (8)* $\forall\, I_t^N$, *under the constraints*

$$\hat{x}_{i+1,t} = f(\hat{x}_{it}, u_i), \quad i = t-N, \ldots, t-1 \tag{19}$$

The minimizations are linked sequentially by the optimal predictions

$$\bar{x}^\circ_{t-N} = f(\hat{x}^\circ_{t-N-1,t-1}, u_{t-N-1}), \qquad t = N+1, N+2, \ldots; \ \bar{x}_0 \in \mathcal{X}$$

□

The initial predictions take values in the rather large set $\mathcal{X}$. Clearly this may be conservative in the light of the minimax procedure described below (at the price of rather involved technical conditions, a more accurate result can be found in [5]).

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*

$$\min_w \max_{\bar{x}_0 \in \mathcal{X}, y_0^N \in Y^{N+1}, r_0^{N-1} \in R^N} \| a^\circ(\bar{x}_0, y_0^N, r_0^{N-1}) - \hat{a}^{(\nu)}(\bar{x}_0, y_0^N, r_0^{N-1}, w) \| \le \bar{\varepsilon} \tag{20}$$

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 (20) is less than or equal to $\bar{\varepsilon}$.

Now, denote by k_a the Lipshitz constant of the optimal estimation function a° and let $\varepsilon \triangleq (1 + k_a k_f)\bar{\varepsilon}$. Then, the following theorem can be proved.

Theorem 3. [5] *Suppose that Assumptions A1) to A6) are verified. Denote by* $\hat{e}_{t-N} \triangleq x_{t-N} - \hat{x}_{t-N,t}$ *the estimation error at stage* $t-N$. *Consider the largest closed ball* $N(\hat{r}_e)$ *with radius* $\hat{r}_e$ *and center in the origin such that* $\hat{e}_0 \in N(\hat{r}_e)$. *If there exists a choice of* μ, *for which the inequalities*

$$(1-k_f)^2 \mu^3 + \left[(3 + k_f^2 - 4k_f)\delta^2 + 4k_f^2 \bar{k}\tilde{k}(k_f - 5)\varepsilon\right] \mu^2 + \left(3\delta^4 - 2k_f\delta^4 - 8k_f^3 \bar{k}\tilde{k}^2 r_\eta - 20k_f^2 \bar{k}\tilde{k}\delta^2\varepsilon\right)\mu + \delta^6 > 0$$

$$(k_f - 1)\mu < \delta^2$$

are satisfied, the second-order equation

$$\begin{aligned}
&\left(2k_f^3 \bar{k}\tilde{k}\mu^2\right)\hat{\xi}^2 + \Big[k_f\mu(\mu+\delta^2)^2 - (\mu+\delta^2)^3 + 4k_f^3\bar{k}\tilde{k}^2\mu r_\eta \\
&\quad + \left(10k_f^2\bar{k}\tilde{k}\mu(\mu+\delta^2) + +4k_f^3\bar{k}\tilde{k}\mu^2\right)\varepsilon\Big]\hat{\xi} \\
&\quad + 2k_f^3\bar{k}\tilde{k}^3 r_\eta^2 + k_f\tilde{k}(\mu+\delta^2)^2 r_\eta \\
&\quad + \left[\frac{25}{4}\bar{k}\tilde{k}(\mu+\delta^2)^3 + 2k_f^3\bar{k}\tilde{k}\mu^2 + 10k_f^2\bar{k}\tilde{k}\mu(\mu+\delta^2)\right]\varepsilon^2 \\
&\quad + \left[10k_f^2\bar{k}\tilde{k}^2(\mu+\delta^2) + 4k_f^3\bar{k}\tilde{k}^2\mu\right]\varepsilon r_\eta \\
&\quad + \left[(\mu+\delta^2)^3 + k_f\mu(\mu+\delta^2)\right]\varepsilon = 0
\end{aligned}$$

has the two real positive roots $\hat{\xi}^-$ *and* $\hat{\xi}^+$, *with* $\hat{\xi}^- < \hat{\xi}^+$. *Then, if the choice of* μ *yields also the fulfilment of the inequality*

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

we have

$$\lim_{t\to+\infty} \|\hat{e}_t\| \le \hat{\xi}^-, \quad \forall\, \hat{r}_e < \hat{\xi}^+ \tag{21}$$

□

To sum up, thanks to Theorem 3, for a given level of measurement noise, and a bound on the initial estimation error, it is possible to check whether an upper bound $\bar{\varepsilon}$ on the approximation error due to the use of the neural approximator do exist, such that the conditions stated in the theorem are fulfilled. In the affirmative, the minimax learning method can be performed, thus giving rise to an approximate nonlinear state estimator characterized by a bounded estimation error.

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, 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, "Neural approximators for nonlinear sliding-window state observers," *Proc. 35th CDC*, Kobe, Japan, 1996.

[4] K. Hornik, M. Stinchombe, and H. White, "Multilayer feedforward networks are universal approximators," *Neural Networks*, vol. 2, pp. 359-366, 1989.

[5] A. Alessandri, T. Parisini, and R. Zoppoli, "Neural networks for nonlinear state estimators," DIST Tech. Report No. 97/1, 1997.