DESIGN OF OBSERVERS FOR CONTINUOUS-TIME NONLINEAR SYSTEMS USING NEURAL NETWORKS

A. Alessandri*, C. Cervellera*, A. F. Grassia*, M. Sanguineti**

* Institute of Intelligent Systems for Automation (ISSIA-CNR)
National Research Council of Italy
Via De Marini 6, 16149 Genova, Italy
{angelo,cervellera,filippo}@ge.issia.cnr.it

** Department of Communications, Computer and System Sciences (DIST)
University of Genoa
Via Opera Pia 13, 16145 Genova, Italy
marcello@dist.unige.it

Abstract— **Observers design is addressed for a class of continuous-time, nonlinear dynamic systems with Lipschitz nonlinearities. A full-order state estimator is considered that depends on an innovation function made up of two terms: a linear gain and a feedforward neural network that provides a nonlinear contribution. The gain and the weights of the neural network are chosen in such way to ensure the convergence of the estimation error. Such a goal is achieved by constraining the derivative of a Lyapunov function to be negative definite on a sampling grid of points. Under assumptions on the smoothness of the Lyapunov function and of the distribution of the sampling points, the negative definiteness of the derivative of the Lyapunov function is obtained by minimizing a cost function that penalizes the constraints that are not satisfied. Suitable sampling techniques allow to reduce the computational burden required by the network's weights optimization. Simulations results are presented to illustrate the effectiveness of the proposed method.**

I. INTRODUCTION

Various methods are reported in the literature to construct observers. A Lyapunov function is usually considered to guarantee the convergence to zero of the estimation error. Unfortunately, there is no general methodology to find such Lyapunov function.

The first convergence results on observers for nonlinear systems were presented in [1] and [2]. The state-space transformation approach (see [3], [4]) allows one to easily find an observer with linear error dynamics in the transformed state-space. In this context, high-gain observers were considered [5] and frequently used in cascade with a regulator for output feedback control. The design of constant-gain observers has been faced in [6], [7]. To deal with the problem of uncertainties, variable-structure observers were proposed [8]. In [9], [10] the design of sliding-mode observers for nonlinear systems was addressed.

In this paper, state estimation problems are considered for continuous-time, nonlinear dynamic systems with Lipschitz nonlinearities by means of a full-order observer that is constructed using a suitable innovation function. Under some regularity assumptions on the system and measurement equations and on the innovation function, a procedure is developed to design such an observer, which can be implemented by means of a class of approximating networks such as feedforward neural networks.

In order to guarantee the convergence of the estimation error, a quadratic Lyapunov function is sought for a parameterized innovation function that is made up of two terms: a linear gain and a feedforward neural network that provides a nonlinear contribution. The design parameters (i.e., the linear gain and the weights of the neural network) can be chosen in such a way to constrain the derivative of a quadratic Lyapunov function to be negative on a sampling grid of points on the Cartesian product of the state space and the estimation error space. This is accomplished by minimizing a cost function that penalizes the constraints that are not satisfied in correspondence of the sampling points. It is worth noting that the selection of the design parameters is made completely off line (see also [11]). This is the main advantage with respect to neural approaches to estimation for nonlinear systems (see, among the most recent ones, [12], [13], [14]) that rely on the on-line adaptation of the neural weights.

Under assumptions on the distribution of the sampling points and smoothness of the Lyapunov function, the negative definiteness of the Lyapunov function's derivative is ensured, thus the resulting observers provides a convergent estimation error. In particular, it is shown that convergence is obtained by using special deterministic sequences that aim at optimizing the dispersion of the sampling points (a measure that quantifies "how uniformly" the points are spread).

The paper is structured as follows. Section II is devoted to the description of the basic assumptions on the class of nonlinear systems considered. The proposed observer is constructed using approximating networks described in Section III. Section IV presents a design method for such

observer. Simulation results are shown in Section V. Brief comments and the conclusions are given in Section VI.

II. System description and basic assumptions

We consider a class of systems described by

$$\begin{cases} \dot{x} = A\,x + f\,(x) \\ y = C\,x\,, \end{cases} \tag{1}$$

where $x \in X \subset \mathbb{R}^n$ is the state vector, X is compact, $y \in \mathbb{R}^m$ is a vector of measurement, $A \in \mathbb{R}^{n\times n}$ and $C \in \mathbb{R}^{m\times n}$ are matrices, and $f : X \to \mathbb{R}^n$. We make the following assumption.

Assumption 1: Let $B_x \stackrel{\triangle}{=} \{x \in X : \|x\| < \bar{x}\,,\, \bar{x} > 0\}$. *Then,*

(i) $f : X \to \mathbb{R}^n$ *is Lipschitz in* B_x*, i.e, there exist* $L_f^x \in \mathbb{R}^+$ *such that* $\|f(x_1) - f(x_2)\| \le L_f^x \|x_1 - x_2\|$*, for all* $x_1, x_2 \in B_x$*;*
(ii) the pair (A, C) *is observable.*

□

Assumption 1 (i) guarantees the existence and uniqueness of a local solution of the differential equation describing the dynamics of the system (1) (see, for example, [15]). Assumption 1 (ii) says that we focus on observable systems that have the state-space representation (1) or that are diffeomorphic to (1) (see, for an introduction, [16]).

A full-order state estimator for (1) is given by

$$\dot{\hat{x}} = A\,\hat{x} + f(\hat{x}) + L\,(y - C\hat{x}) + \gamma\,(y - C\hat{x})\,, \tag{2}$$

where $L \in \mathbb{R}^{n\times m}$ is a matrix and $\gamma : \mathbb{R}^m \to \mathbb{R}^n$ is a function. The following assumption defines admissible functions γ.

Assumption 2: Let $z \stackrel{\triangle}{=} y - C\hat{x} \in Z \subset \mathbb{R}^m$ *and* $B_z \stackrel{\triangle}{=} \{z \in Z : \|z\| < \bar{z}\,,\, \bar{z} > 0\}$. *Then* $\gamma : Z \to \mathbb{R}^n$ *is locally Lipschitz in* B_z *and such that* $\gamma(0) = 0$. □

The Lipschitz condition in Assumption (2) is a sufficient requirement for having a unique local solution of the differential equation (2) describing the estimator. The condition $\gamma(0) = 0$ in Assumption (2) guarantees that, in the absence of disturbances, if there exists $T \ge 0$ such that $\hat{x}(T) = x(T)$, then $\hat{x}(t) = x(t)$ for every $t \ge T$ (see, e.g., [17]). The innovation function has been chosen with the above-written form for the sake of simplicity, although it can be of a more general type [17], e.g., with a function $\gamma\,(Cx, C\hat{x})$ such that $\gamma(z, z) = 0\,,\ \forall z \in Z$.

III. Parameterized estimators via approximating networks

We face the problem of designing an estimator for system (1) by searching for a matrix L and a function γ associated with a suitable Lyapunov function for the estimation error $e(t) = x(t) - \hat{x}(t)$ of the observer (2). Towards this end, we further restrict the class (2) of observers by considering functions γ that not only satisfy Assumption (2), but that also have a special structure, corresponding to linear combinations of functions with a fixed structure and depending on a vector $w_\nu \in \mathbb{R}^l$ of parameters. More precisely, for every $\nu \in \mathbb{N}$ we define the following class of functions. By $\mathcal{C}(K, \mathbb{R}^n)$ we denote the space of continuous functions defined on a compact set $K \subset \mathbb{R}^m$, equipped with the supremum norm.

Definition 1: Approximating networks of order ν *are functions belonging to the set*
$A_\nu \stackrel{\triangle}{=} \{\ \gamma_\nu : K \times \mathbb{R}^l \to \mathbb{R}^n$ such that

(i) $\gamma_{\nu\,j}(\xi, \omega_{\nu_j}) = \sum_{i=1}^{\nu} c_{ij}\,\varphi_i(\xi, \kappa_i)$, $\varphi_i : K \times \mathbb{R}^l \to \mathbb{R}$, $|c_{ij}| \le A$, $A \in \mathbb{R}^+$, $\kappa_i \in \mathbb{R}^l$, $i = 1, \ldots, \nu$, $j = 1, \ldots, n$, $\omega_{\nu\,j} \stackrel{\triangle}{=} \mathrm{col}(c_{ij}, \kappa_i : i = 1, \ldots, \nu)$;
(ii) $\varphi_i(\cdot, \kappa_i)$ bounded in aggregate, i.e., $\exists\, M \in \mathbb{R}^+$ such that $\forall i = 1, \ldots, \nu$, $\forall \kappa_i \in \mathbb{R}^l$, $\sup_{\xi\in K} |\varphi_i(\xi, \kappa_i)| \le M$;
(iii) $\varphi_i(\cdot, \kappa_i)$ is Lipschitz, i.e., $\forall i = 1, \ldots, \nu$ $\exists\, L_i \in \mathbb{R}^+$ such that $\forall \kappa_i \in \mathbb{R}^l$, $|\varphi_i(\xi, \kappa_i) - \varphi_i(\xi', \kappa_i)| \le L_i |\xi - \xi'|$*;*
(iv) $\varphi_i(\cdot, \kappa_i) = 0\,,\ i = 1, \ldots, \nu\,,\ \forall \kappa_i \in \mathbb{R}^l\,.\ \}$.

□

Note that items (iii) and (iv) in Definition 1 take into account the requirements of Assumption 2.

Thus, the estimator (2) takes on the form

$$\dot{\hat{x}} = A\,\hat{x} + f(\hat{x}) + L\,(y - C\hat{x}) + \gamma_\nu\,(y - C\hat{x}, w_\nu)\,, \tag{3}$$

where $\gamma_\nu : \mathbb{R}^n \times \mathbb{R}^l \to \mathbb{R}^n \in A_\nu$.

The definition above characterizes approximating networks as a particular kind of linear combinations of variable-basis functions. More precisely, for each i $\varphi_i(\cdot, \cdot)$ is a given basis function and c_{ij} and the components of κ_i are free parameters, lumped together in the vector $w_\nu \stackrel{\triangle}{=} \mathrm{col}\,(w_{\nu\,j}\,, j = 1, \ldots, n) \in \mathbb{R}^{\mathcal{N}(\nu)}$, where $\mathcal{N}(\nu) = mk\nu + n\nu$. In practice, as the inner parameters $\kappa_i$in each basis function allow a wide flexibility, typically variable-basis functions are obtained by varying such parameters in a unique "mother function" $\varphi(\cdot, \cdot)$. The properties of variable-basis approximation and their application to optimization problems have been extensively studied in [18], [19], [20], [21].

As each γ_ν is a sum of ν Lipschitz functions bounded by $A\,M$, it is Lipschitz. Then, we have the following

Proposition 1: For each ν*, approximating Networks of order* ν *are admissible innovation functions.* □

Once a type of approximating networks, i.e., a mother function $\varphi(\cdot, \cdot)$ is chosen, the Lyapunov function for the estimator (3) depends on the values of L and w_ν. As to the function φ, it is suitable to make a choice generating

sets A_ν that have a closure as large as possible: loosely speaking, the larger such a closure, the wider the choice at our disposal for a Lyapunov function. So we shall use approximating networks that are dense in the space of continuous functions on compact sets, i.e., in the neural-network parlance, that enjoy the "universal approximation property". Well-known examples of approximating networks are feedforward neural networks of the perceptron type, with at most ν hidden units and bounded parameters and radial-basis-functions with at most ν hidden units and bounded input weights and variances. The proofs of the fact that such functions are provided with the density property in the space $\mathcal{C}(K, \mathbb{R}^n)$, where $K \subset \mathbb{R}^m$ is compact, can be found, for example, in [22] and [23].

To guarantee the possibility of finding an estimator implemented with a "small" number ν of basis functions also for vectors x to be estimated with a large number of components, we shall employ so-called "polynomially-complex approximating networks". Such networks have the desirable property that the number ν of basis functions required to guarantee a fixed approximation accuracy has to grow, under mild conditions, at most polynomially with the number of variables (in the case of the estimator (3), the dimension m of the measurement vector); see [18], [19], [24] for details.

Once the structure of the approximating network γ_ν is fixed, the parameters L and w_ν are chosen, and the output y is known, the evolution of the estimated state vector $\hat{x}$ is completely determined by (3) according to the intial conditions. Of course, the selection of L and w_ν must ensure the stability of the estimation error.

By (1) and (3), the dynamics of the estimation error is given by

$$\dot{e}= (A - LC)\, e + f(x) - f(\hat{x}) - \gamma_\nu\,(C\, e, w_\nu) \quad . \quad (4)$$

If a quadratic Lyapunov function $V = e^{\mathrm{T}}\, P\, e$ is considered with a symmetric positive definite matrix P, we obtain

$$\begin{aligned} \dot{V} &= e^{\mathrm{T}}\, \left[(A - LC)^{\mathrm{T}}\, P + P\,(A - LC)\right]\, e \\ &+ 2\, [f(x) - f(\hat{x}) - \gamma_\nu\,(C\, e, w_\nu)]^{\mathrm{T}}\, P\, e \quad . \end{aligned}$$

By Assumption 1 (ii) the pair (A, C) is observable, so, for any gain matrix L such that $A - LC$ is Hurwitz and any symmetric positive definite matrix Q, there exists a unique symmetric positive definite matrix P solving the Lyapunov equation

$$(A - LC)^{\mathrm{T}}\, P + P\,(A - LC) = -Q\ .$$

Therefore, in order to guarantee the asymptotic stability of the estimation error, we can impose $\dot{V} \leq 0$, i.e.,

$$2\, [f(x) - f(x - e) - \gamma_\nu\,(C\, e, w_\nu)]^{\mathrm{T}}\, P\, e - e^{\mathrm{T}}\, Q\, e \leq 0$$

along the trajectories of both the system and the estimation error dynamics. In the next section, we shall address this specific issue.

IV. Design of the parameterized estimator by sampling sequences and optimization

Let $E \subset \mathbb{R}^n$ be a compact set where the estimation error takes values and define $S_M \subset X \times E$ as a set of M sample points $s_i \stackrel{\triangle}{=} (x_i, e_i)^{\mathrm{T}}$ for $i = 1, \ldots, M$ in the Cartesian product of the state and estimation error spaces and consider

$$\begin{aligned} \dot{V}\,(s_i, w_\nu) \stackrel{\triangle}{=} 2\,[f(x_i) - f(x_i - e_i) \\ - \quad \gamma_\nu\,(C\, e_i, w_\nu)]^{\mathrm{T}}\, P\, e_i - e_i^{\mathrm{T}}\, Q\, e_i\ , \\ i = 1, 2, \ldots, M\ . \quad (5) \end{aligned}$$

In the following, we show how it is possible to prove the asymptotic convergence of the estimation error by imposing the negativity of the derivative $\dot{V}\,(s_i, w_\nu)$ of the Lyapunov function on the points of S_M. This corresponds to finding a w_ν (i.e., optimizing a suitable approximating network γ_ν) such that

$$\dot{V}\,(s_i, w_\nu) \leq 0\ ,\ i = 1, 2, \ldots, M. \quad (6)$$

In order to ensure that the derivative $\dot{V}\ (s, w_\nu)$ is negative for *any* s that belongs to a compact set $S \stackrel{\triangle}{=} E \times X$, we exploit the regularity of $\dot{V}$ and choose a sampling rule that guarantees a suitably dense covering of S.

To this end, define $w^*_{\nu,M}$ as the parameters vector obtained after optimizing the network γ_ν over M sample points, and let $\dot{V}^*(s) \stackrel{\triangle}{=} \dot{V}(s, w^*_{\nu,M})$. By the Lipschitz continuity of the functions f and γ, it follows that for any $s, s' \in S$ the function $\dot{V}^*$ is Lipschitz too, i.e., there exists $L_V \in \mathbb{R}^+$ such that $|\dot{V}^*(s) - \dot{V}^*(s')| \leq L_V \|s - s'\|$, .

As $\dot{V}^*(s_i) < 0$ for all $s_i \in S_M$, $i = 1, \ldots, M$, there exists $\epsilon_V > 0$ such that

$$\epsilon_V \stackrel{\triangle}{=} - \max_{1 \leq i \leq M} \dot{V}^*(s_i)\ . \quad (7)$$

Then, for any $s \in S$, we can write

$$\dot{V}^*(s) \leq \dot{V}^*(\tilde{s}) + L_V \|s - \tilde{s}\| \leq L_V \sup_{s \in S} \|s - \tilde{s}\| - \epsilon_V$$

where $\tilde{s} \stackrel{\triangle}{=} \arg \min_{1 \leq i \leq M} \|s - s_i\|$ is the sample point closest to s. The quantity $\theta\,(S_M) \stackrel{\triangle}{=} \sup_{s \in S} \min_{1 \leq i \leq M} \|s - s_i\|$ is called the *dispersion* of the sequence of M points [25] and is a measure of their uniformity of distribution. Thus

$$\dot{V}^*(s) \leq L_V \theta\,(S_M) - \epsilon_V\ ,$$

So, $\dot{V}^*(s)$ is negative definite provided that the sample sequence and the number of points M are such that

$$\theta\,(S_M) < \frac{\epsilon_V}{L_V}$$

Therefore, by [15][Theorem 3.1, p. 100], the estimation error converges to zero.

In other words, we must guarantee that (i) the points of S_M are spread in the most uniform way on S (in such a way that $\theta(S_M)$ is small), without leaving regions of the

space "undersampled" and (ii) that the points are "close enough" to each other.

The discussion above enables us to state the following result.

Theorem 1: Consider observer (3) for system (1) and suppose that Assumptions 1 and 2 hold. If there exist a gain matrix L, two symmetric positive definite matrices P and Q, a set $S_M \subset E \times X$ of sampling points, and a parameters vector $w^*_{\nu,M}$ such that

$$(A - LC)^{\mathrm{T}} P + P(A - LC) = -Q \tag{8}$$

$$\dot{V}\,(s_i, w^*_{\nu,M}) \leq 0\ ,\ s_i \in S_M, \quad i = 1, 2, \ldots, M\ , \tag{9}$$

$$\theta\,(S_M) < \frac{\epsilon_V}{L_V}\ , \tag{10}$$

then for any $e(0) \in E$ the estimation error of the observer (3) converges to zero.

□

In order to satisfy the constraints expressed by (9), we need to choose an approximating network γ_ν of a sufficiently large order and optimize its parameters by minimizing a cost function that penalizes non-satisfied constraints. Following [26], we use the cost

$$J = \sum_{i=1}^{M} \left(\max \left\{ 0, \dot{V}\,(s_i, w_{\nu,M}) \right\} \right)^2\ . \tag{11}$$

In the particular case in which the approximating networks γ_ν are given by neural networks, the way of optimizing their parameters is quite different from standard neural-network learning, where the goal is to minimize the distance of the network's output from given target values. In our approach, as done in [26], we use an optimization sometimes referred to as *distal training*; so we have to employ ad-hoc techniques, which are often modified versions of standard minimization algorithms. For example, in [26] the satisfaction of the constraints is obtained by minimizing a suitable quadratic penalty function using a specialized version of the Levenberg-Marquardt algorithm. Note also that this approach is somehow dual to the approach reported in [26] to design closed-loop neural controllers for nonlinear systems. For our simulations, we used a Monte Carlo technique combined with a clustering algorithm, typically employed in global optimization problems (see [27], [28]).

Condition (10) is related to the dispersion properties of the grid S_M. Monte Carlo sampling with uniform distribution can satisfy these requirements, as long as the sample size M is large enough. The best dispersion properties, in the sense described above, belong to special *deterministic* sequences called *low-discrepancy sequences*, commonly employed in the fields of number-theoretic methods, statistics and quasi Monte Carlo integration. Examples of such sequences are $(t, n) - sequences$, the *Halton sequence*, and the *Hammersley sequence* [27], [25]. The use of low-discrepancy sequences for function learning by neural networks is described in [29].

All such sequences attain deterministically a rate of convergence for the dispersion of order $O\left(M^{-1/2n}\right) \leq \theta\,(S_M) \leq O\left(\sqrt{2n} M^{-1/2n}\right)$, where the number M of points required to attain a desired dispersion can be computed exactly. With respect to pure Monte Carlo sampling, it is proven that these discretization schemes suffer less from the formation of clusters of points in particular regions of the space, which undermines the uniformity of the sampling.

V. Numerical results

We consider the Van der Pol equation given by

$$\begin{aligned} \dot{x}_1 &= x_2 \\ \dot{x}_2 &= -9x_1 + 2(1 - x_2^2)x_2 \\ y &= x_1 \quad . \end{aligned} \tag{12}$$

Although system (12) is not stable, it admits a limit cycle; so the hypothesis of the compactness of X is satisfied.

The design of an observer for (12) was made by choosing $L = (0.03, -0.06)^{\mathrm{T}}$ and Q equal to the identity and solving the corresponding Lyapunov equation to find P.

The results obtained for $x_1(0) = x_2(0) = 0$ with a selection of the initial estimated values $\hat{x}_1(0) = \hat{x}_2(0) = 3$ by using a 3-neuron feedforward neural network with 100 and 700 sampling points for training are shown in Fig.s 1 and 2, respectively. The results in the same conditions with a 10-neuron feedforward neural network with 100 and 700 sampling points are shown in Fig.s 3 and 4, respectively.

From the figures it turns out that the performances at steady state are quite good for all the cases. The transient behaviors of the estimators depend on both the approximation capability of the neural networks and on the effectiveness of the training. The number of sampling points may considerably affect the performances, as a larger neural network requires a larger set of sampling points. The bad behavior shown in Fig. 3 for an observer with a 10-neuron feedforward neural network trained using only 100 sampling points may be ascribed to an overfitting phenomenon, due to the small number of discretization points with respect to the number of neural units.

VI. Conclusions

A new method to design observers for a class of nonlinear systems has been presented. The observer depends on an innovation function that is made up of two terms: a linear gain and a feedforward neural network that provides a nonlinear contribution. The gain and the neural network weights are chosen in such way to guarantee the convergence to zero of the estimation error by searching for a sufficiently smooth Lyapunov function. We have shown that such a convergence can be obtained by constraining the derivative of the Lyapunov function to be negative on a well-shaped sampling grid of points, as this, together with additional conditions on the linear gain, ensures the asymptotic stability of the estimation error. For what concerns the sampling points, the

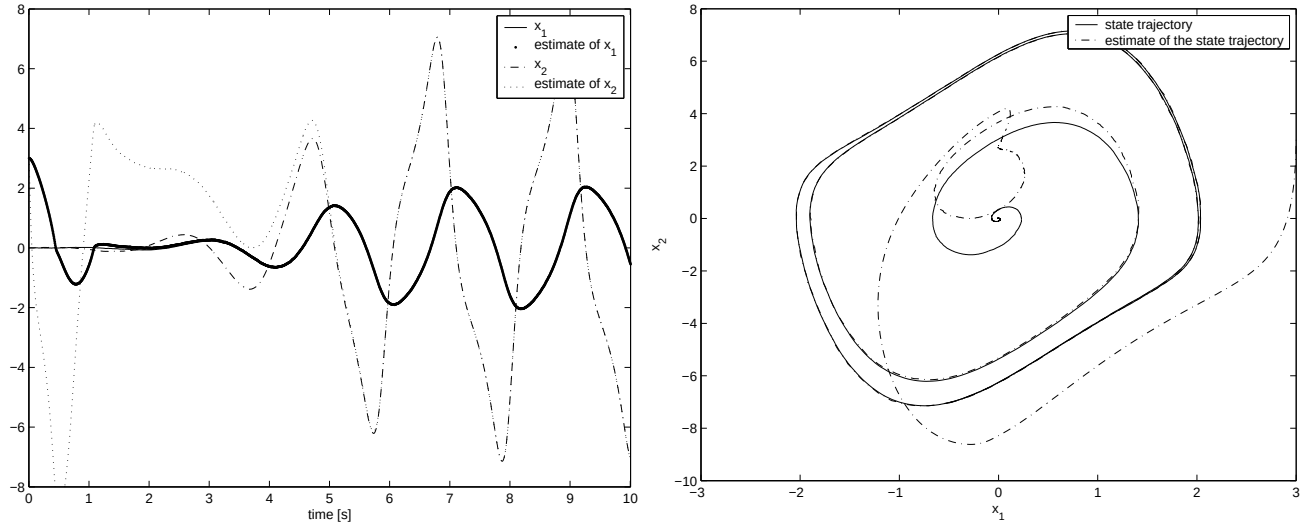

Fig. 1. Numerical results with a 3-neuron feedforward neural network and 100 sampling points for training.

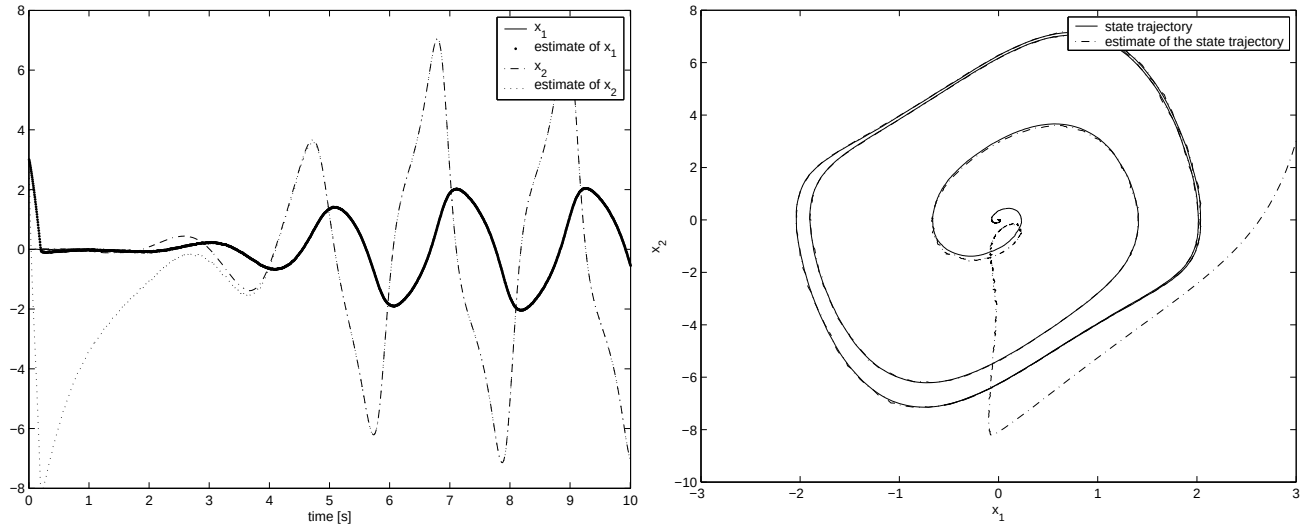

Fig. 2. Numerical results with a 3-neuron feedforward neural network and 700 sampling points for training.

use of so-called low-discrepancy sequences, which ensure deterministic convergence with favorable rates when the number of points grows, has been discussed.

REFERENCES

[1] F. E. Thau, "Observing the state of non-linear dynamic systems", *Int. J. of Control*, vol. 17, no. 3, pp. 471–479, 1973.

[2] S. R. Kou, D. L. Elliott, and T. J. Tarn, "Exponential observers for nonlinear dynamic systems", *Information and Control*, vol. 29, pp. 204–216, 1975.

[3] A. J. Krener and A. Isidori, "Linearization by output injection and nonlinear observers", *Systems & Control Letters*, vol. 3, pp. 47–52, 1983.

[4] A. J. Krener and W. Respondek, "Nonlinear observer with linearizable error dynamics", *SIAM J. on Control and Optimization*, vol. 23, no. 2, pp. 197–216, 1985.

[5] J. P. Gauthier, H. Hammouri, and S. Othman, "A simple observer for nonlinear systems applications to bioreactors", *IEEE Trans. on Automatic Control*, vol. 37, no. 6, pp. 875–880, 1992.

[6] S. Raghavan and J. K. Hedrick, "Observer design for a class of nonlinear systems", *Int. J. of Control*, vol. 59, no. 2, pp. 515–528, 1994.

[7] R. Rajamani, "Observers for Lipschitz nonlinear systems", *IEEE Trans. on Automatic Control*, vol. 43, no. 3, pp. 397–401, 1998.

[8] V. I. Utkin, *Sliding Modes in Control and Optimization*, Springer-Verlag, 1992.

[9] B. L. Walcott and S. H. Zak, "State observation of nonlinear uncertain dynamical systems", *IEEE Trans. on Automatic Control*, vol. 32, no. 2, pp. 166–170, 1987.

[10] E. Yaz and A. Azemi, "Variable structure observer with a boundary-layer for correlated noise/disturbance models and disturbance minimization", *Int. J. of Control*, vol. 57, no. 5, pp. 1191–1206, 1993.

[11] A. Alessandri, M. Baglietto, T. Parisini, and R. Zoppoli, "A neural state estimator with bounded errors for nonlinear systems", *IEEE Trans. on Automatic Control*, vol. 44, no. 11, pp. 2028–2042, 1999.

[12] J. Y. Choi and J. A. Farrel, "Adaptive observer backstepping control using neural networks", *IEEE Trans. on Neural Networks*, vol. 12, no. 5, pp. 1103–1112, 2001.

[13] A. G. Parlos, S. K. Menon, and A. F. Atiya, "An algorithmic approach to adaptive state filtering using recurrent neural networks", *IEEE Trans. on Neural Networks*, vol. 12, no. 6, pp. 1411–1432, 2001.

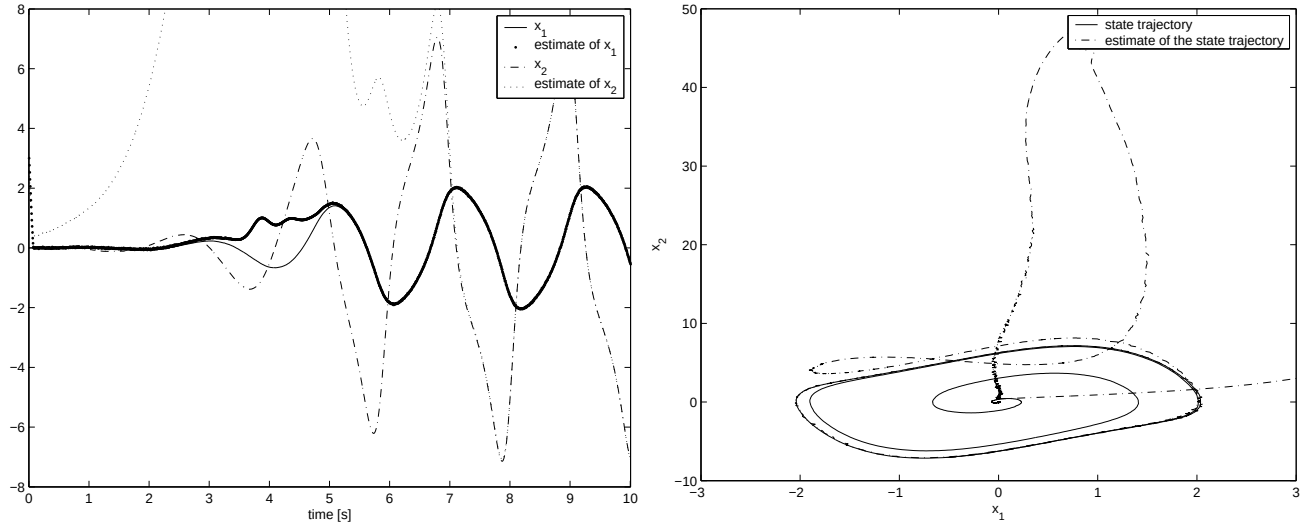

Fig. 3. Numerical results with a 10-neuron feedforward neural network and 100 sampling points for training.

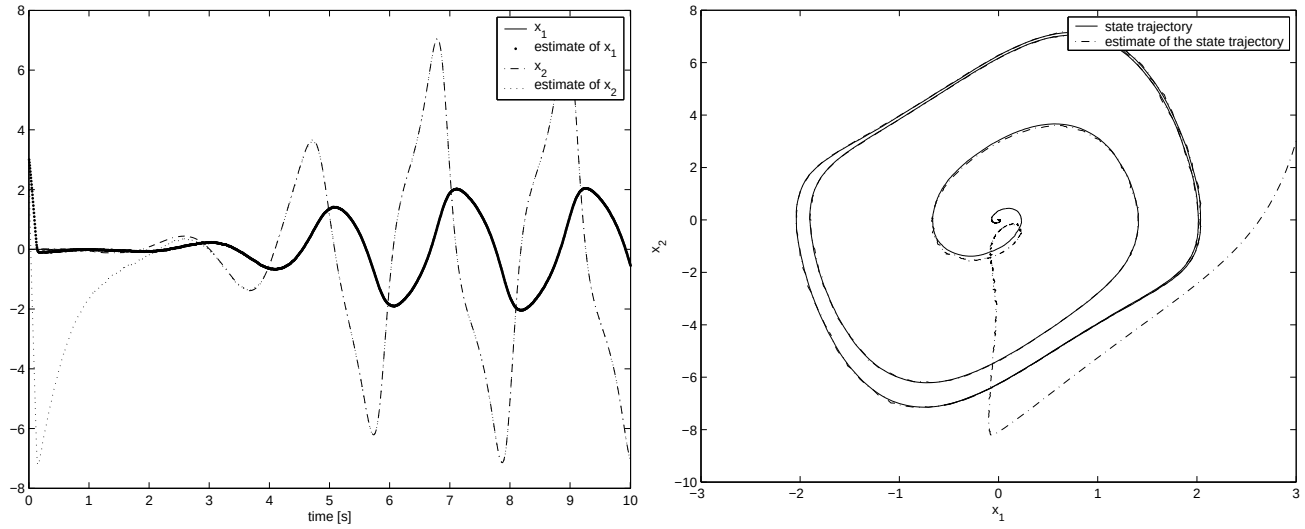

Fig. 4. Numerical results with a 10-neuron feedforward neural network and 700 sampling points for training.

[14] X. Zhang, M. M. Polycarpou, and T. Parisini, "A robust detection and isolation scheme for abrupt and incipient faults in nonlinear systems", *IEEE Trans. on Automatic Control*, vol. 47, no. 4, pp. 576–593, 2002.

[15] H. K. Khalil, *Nonlinear Systems*, Prentice Hall, 1996.

[16] H. Shim, Y. I. Son, and J. H. Seo, "Semi-global observer for multi-output nonlinear systems", *System & Control Letters*, vol. 42, no. 3, pp. 233–244, 2001.

[17] X. Xia and W. Gao, "On exponential observers for nonlinear systems", *Systems & Control Letters*, vol. 11, pp. 319–325, 1988.

[18] V. Kurková and M. Sanguineti, "Bounds on rates of variable-basis and neural-network approximation", *IEEE Trans. on Information Theory*, vol. 47, pp. 2659–2665, 2001.

[19] V. Kurková and M. Sanguineti, "Comparison of worst case errors in linear and neural network approximation", *IEEE Trans. on Information Theory*, vol. 48, pp. 264–275, 2002.

[20] P. Kainen, V. Kurková, and M. Sanguineti, "Minimization of error functionals over variable-basis functions", *SIAM J. on Optimization*, 2004 (to appear).

[21] V. Kurková and M. Sanguineti, "Error estimates for approximate optimization by the extended ritz method", *SIAM J. on Optimization*, 2004 (accepted).

[22] M. Leshno, V. Ya, A. Pinkus, and S. Schocken, "Multilayer feedforward networks with a nonpolynomial activation function can approximate any function", *Neural Networks*, vol. 6, pp. 861–867, 1993.

[23] V. Kůrková, "Approximation of functions by perceptron networks with bounded number of hidden units", *Neural Networks*, vol. 8, pp. 745–750, 1995.

[24] R. Zoppoli, M. Sanguineti, and T. Parisini, "Approximating networks and extended Ritz method for the solution of functional optimization problems", *J. of Optimization Theory and Applications*, vol. 112, no. 2, 2002.

[25] H. Niederreiter, *Random Number Generation and Quasi-Monte Carlo Methods*, SIAM, Philadelphia, 1992.

[26] S. Yu and A. M. Annaswamy, "Stable neural controllers for nonlinear dynamic systems", *Automatica*, vol. 34, no. 5, pp. 641–650, 1998.

[27] K.-T. Fang and Y. Wang, *Number-theoretic Methods in Statistics*, Chapman & Hall, London, 1994.

[28] A. Torn and A. Zilinskas, *Global Optimization*, Springer-Verlag, New York, 1989.

[29] C. Cervellera and M. Muselli, "Deterministic design for neural network learning: an approach based on discrepancy", *IEEE Trans. on Neural Networks*, 2004 (in press).